2017

Metabolic engineering strategies for high-level production of aromatic amino acid pathway derivatives in Saccharomyces cerevisiae

Jose Miguel Suastegui Pastrana
Iowa State University

Metabolic engineering strategies for high-level production of aromatic amino acid pathway derivatives in *Saccharomyces cerevisiae*

by

José Miguel Suástegui Pastrana

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Chemical Engineering

Program of Study Committee:
Zengyi Shao, Major Professor
Julie A. Dickerson
Laura R. Jarboe
Basil J. Nikolau
Brent H. Shanks

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

NOMENCLATURE

ARO1	Pentafunctional gene in the shikimic acid pathway
BIAs	Benzylisoquinoline alkaloids
CSM	Complete Synthetic Medium
E4P	Erythrose-4-Phosphate
FBA	Flux Balance Analysis
HIS	Histidine
L-PHE	L-Phenylalanine
L-TRP	L-Tryptophan
L-TYR	L-Tyrosine
LEU	Leucine
MA	Muconic Acid
MFA	Metabolic Flux Analysis
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
OAA	Oxaloacetate
PCA	Protocatechuic Acid
PEP	Phosphoenolpyruvate
PPP	Pentose Phosphate Pathway
SA	Shikimic Acid
URA	Uracil
YPAD	Yeast extract – Peptone – Adenine – Dextrose medium

ACKNOWLEDGMENTS

First, I would like to thank my thesis advisor, Dr. Zengyi Shao for her continuous support during my doctoral studies. Her guidance and mentorship have encouraged me to strive for excellence in research. I feel blessed for having had the opportunity of being her first Ph.D. student. I would also like to thank my committee members Dr. Basil Nikolau, Dr. Brent Shanks, Dr. Laura Jarboe, and Dr. Julie Dickerson, for their support and for providing invaluable advice for the preparation of this thesis.

I would like to show my gratitude for the members of the Center for Biorenewable Chemicals (CBiRC). Besides being a major source of funding for my research, this center provided me with opportunities to establish strong and very fruitful collaborations with researchers outside my area of expertise. I want to especially thank John Matthiesen and his advisor Dr. Jean-Phillippe Tessonnier for sharing their passion and dedication in every step of this journey.

CBiRC also gave me the opportunity to participate in leadership roles through the Student Leadership Council. I want to thank all my SLC friends, especially Kayla Flyckt who was my partner in crime as co-chair, and Dr. Alexis Campbell, for always providing such invaluable professional and life advice. Additionally, I would like to thank all the members of the Shao group: Mingfeng, Meirong, Le, Wan, Carmen, Aric, Yutong, Deon, and Yanzhen for their constant encouragement and support. Especial thanks go to the all the bright minds on the 4th floor of the Biorenewables Research Lab for making my time at Iowa State University a wonderful experience.

Finally, and most importantly, I want to thank my parents and my sister, without whom none of this could have ever been possible. Their unconditional love and support have provided me with the valor to pursue my dreams, and for that, I am eternally grateful.

ABSTRACT

Due to its robustness, genetic tractability, and industrial relevance, the budding yeast *Saccharomyces cerevisiae* was selected as study model of the aromatic amino acid biosynthetic pathway. This pathway houses a wide diversity of economically important metabolites ranging from polymer precursors to pain-management drugs, whose productions have been highly sought-after in biotechnological research. However, tight regulations at the transcriptional, translational, and allosteric levels surround the aromatic amino acid pathway, protecting the microbial factories (*e.g. S. cerevisiae*) from unnecessary energy expenditures. By making use of computational metabolic engineering tools such as Flux Balance Analysis and Metabolic Flux Analysis, together with fast and reliable synthetic biology techniques, the flux into the aromatic amino acid pathway was exploited. Initially, the flux distribution in the central carbon metabolism was studied through ^{13}C-metabolic flux analysis and carbon tracing experiments. Important insights regarding the partition between glycolysis and the pentose phosphate pathway were obtained and correlated with the production of aromatic amino acid derivatives. For the first time, the pentafunctional enzyme, ARO1, composing the core of the shikimic acid pathway was subjected to site-directed mutagenesis to reveal its active domains. This resulted in the development of new variants with disrupted activities specifically designed for increasing production of the two target molecules, namely, muconic acid and shikimic acid. Further analysis with OptForce simulations revealed that overexpressing the ribose-5-phosphate ketol-isomerase gene, *RKI1*, can enhance carbon funneling into the aromatic amino acid pathway. A multilevel engineering strategy was established to explore novel transcriptional regulators that tightly control the carbon flux into the pathway. Deleting the

gene *RIC1*, involved in efficient protein localization of trans-Golgi network proteins, increased the titers of shikimic acid and muconic acid. These non-intuitive interventions, in combination with the previous genetic platforms, increased the production titers over 3-fold compared to the base strains. The shikimic acid strains produced 1.9 g L-1, while muconic acid and intermediates were accumulated up to 1.6 g L-1, both being the highest reported in *S. cerevisiae*, in batch fermentations. Future research should focus on devising more dynamic genome engineering strategies that rely on modulating the activity of essential genes while ensuring a good compromise with biomass formation.

CHAPTER 1 INTRODUCTION

1. Motivation

The environmental impact caused by exploiting the natural sources of fossil fuels has led to extensive research in the biotechnology industry. The fermentation of cheap carbon sources through engineered microorganisms has proven to constitute a promising alternative to classic chemical synthesis processes that otherwise employ petrochemical feedstocks. For example, the manipulation of microorganisms like *Escherichia coli* and *Saccharomyces cerevisiae* has rendered powerful microbial factories for the production of a wide array of chemicals with diverse applications. Potentially, this reduces the exploitation of non-renewable fossil resources, and for several specialty chemicals, it removes the dependency of cumbersome extraction processes from plant tissues. In this light, fermentation with metabolically engineered microorganisms stands as a strong alternative with great potential to contribute towards diminishing worldwide economic and environmental issues.

Utilizing microbes for the production of chemicals is an endeavor with decades of intense research. Since the conception of metabolic engineering over three decades ago, the budding yeast *Saccharomyces cerevisiae* has been an exceptional model organism for the discovery and conceptualization of entire production routes. Nowadays, with the development of *state-of-the-art* technology, it is possible to model the yeast's physiology from a molecular to a system-wide level. Starting from the genetic landscape, whole genome sequencing, as well as powerful techniques like RNA-seq, have allowed exploiting of the concept of *synthetic biology*. Furthermore, knowing the genetic makeup of the cell permits its reconstruction into metabolic pathways for mathematical modeling of fluxes. The techniques denominated Flux Balance Analysis (FBA) and Metabolic Flux Analysis (MFA) are key to perform *in silico* simulations

for predictions of complicated metabolic states. Moreover, recent technologies for genome engineering have opened the door to unprecedented advances in the establishment of industrial-scale microbial platforms. For example, the discovery and repurposing of the prokaryotic adaptive immune system, composed of Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes (CRISPR-Cas), has unlocked a whole new era of powerful gene-editing tools.

The system-wide concept of metabolic engineering has been put into practice with exemplary results. Some success stories in *S. cerevisiae* include the establishment of metabolic routes that challenge and revolutionize the concept of simple alcoholic fermentation. One great example is the production of the antimalarial drug artemisinin, and most recently the *de novo* production of the aromatic alkaloids. In this last case, however, it was clear that despite the intensive metabolic engineering applied, the yields still remained far from industrial-scale consideration.

The family of aromatic amino acid derivatives is of particular interest to many sectors of the chemical industry. It encompasses a millionaire market including molecules that range from polymer precursors (*e.g.* muconic acid), drug precursors (*e.g*. shikimic acid), nutraceuticals (*e.g*. naringenin and resveratrol), and potent alkaloids (*e.g*. morphine). Yet, the scenarios of low product yields have evidenced a need to approach the engineering from different and more innovative perspectives. In this light, this thesis contains a comprehensive set of studies that combines the implementation and discovery of new metabolic engineering approaches towards establishing yeast microbial platforms for the biosynthesis of compounds derived from the aromatic amino acid pathway. Such approaches can be repurposed and expanded for other chemical derivatives of industrial importance.

2. Objectives

The central goals supporting the core of this thesis were accomplished through the establishment of the following key objectives:

- Evaluate the metabolic impact and connectivity of the aromatic amino acid pathway in relationship to the key pathways composing the central carbon metabolism in *S. cerevisiae*
    - Combine computational simulation with experimental calculations to predict carbon distributions among glycolysis, pentose phosphate pathway, and the aromatic amino acid pathway
    - Determine the minimal genetic elements required to engineer the central carbon metabolism to direct flux into production of aromatic chemicals
    - Determine the most suitable yeast expression host to enhance production titers
- Establish multilevel engineering strategies to rationally engineer the production of aromatic amino acid derivatives
    - Identify and engineer endogenous enzymatic catalysts to allow carbon channeling for desired products
    - Identify transcriptional regulators that specifically target the family of genes in the aromatic amino acid pathway
    - Identify new metabolic interventions through implementing *in silico* systems-wide genetic reconstruction models

3. Approach

To enhance product formation deriving from the aromatic amino acid pathway in *S. cerevisiae*, computational and experimental metabolic engineering tools were implemented. It is important to mention that although the application for production of aromatic derivatives is case specific, the approach for pathway visualization and engineering can be adopted for any other product deriving from the same pathway or from a distinct one. Initially, the approach consisted of partitioning the aromatic pathway in different modules to localize the perturbations of interest. This allowed directing the attention toward targeting specific bottlenecks for a step-by-step resolution of the overall pathway. New versions of the pentafunctional enzyme Aro1 were developed to enable the accumulation of target compounds. To explore novel genetic interventions, *in silico* simulations of gene perturbations were employed through Metabolic Flux Analysis in combination with Flux Balance Analysis. Shikimic acid served as the molecular marker to track the titer improvements achieved by implementing the novel engineering strategies. By establishing a minimal genetic pathway for shikimic acid accumulation, a strain-dependent approach for metabolic engineering was successfully devised. This approach provides important insights into key regulatory points in the pathway of interest from comparing wildtype strains belonging to the same species but with slightly different genetic makeups. Moreover, this minimal genetic platform permitted the exploration of transcriptional regulator targets that negatively impact the production of aromatic compounds.

Collectively, these strategies compile a series of efforts that led to the construction of competitive aromatic derivative production strains. A direct application of this work encompassed expanding the aromatic pathway toward the production of a heterologous

compound, namely muconic acid. Furthermore, a full production cycle was demonstrated where coupling the biological- and electrocatalysis allowed closing a common gap observed in the efforts of converting cheap sugars to industrially relevant products.

The manipulations presented here for the production of aromatic derivatives have also been adopted for the production of similar compounds in the non-conventional yeast *Pichia stipitis* which possesses higher yield capabilities. This demonstrates that the innovative engineering rationale established in this thesis can serve as an engine for knowledge generation and can be easily be adapted for establishing microbial factories with industrial capabilities beyond the model organism *S. cerevisiae*.

4. Report Organization

An introduction to the aromatic amino acid biosynthetic pathway is presented in Chapter 2. This comprehensive review introduces several aromatic compounds, their economic importance, and their industrial application. Moreover, it covers the most recent metabolic engineering strategies to enable the production of aromatic compounds in *S. cerevisiae*. Next, Chapter 3 is devoted to constructing the highest producing strain of muconic acid reported in the literature. This project was done in collaboration with the Tessonnier group and the Cochran group (ISU, Chemical Engineering) for the diversification of muconic acid for applications in the polymer industry. Chapter 4 contains the first evidence of high production of shikimic acid in yeast and presents a minimal genetic approach to engineering the carbon entrance into the aromatic amino acid pathway. Moreover, in this chapter, the carbon flux into the aromatic pathway was studied through ^{13}C MFA and carbon tracing. Chapter 5 is dedicated to fully engineering the precursor module of the pathway in a multilevel approach. This

included the discovery of new genetic targets for deletion and overexpression from OptForce simulations, as well as one of the firsts attempts to unravel transcriptional regulations in the pathway. Chapter 6 describes several new strategies toward rewiring gluconeogenesis to enhance the production of aromatic compounds. Furthermore, it includes preliminary work on engineering *S. cerevisiae* and *P. stipitis* for the production of plant secondary metabolites. The conclusions of this thesis and future perspectives are also included in this final chapter.

Lastly, synthetic biology techniques for DNA assembly and optimization of pathways were highly used throughout this thesis. Hence, an appendix is included at the end of this document where synthetic biology techniques and computation tools are described to serve as a reference for the construction and optimization of any metabolic pathway.

CHAPTER 2

YEAST FACTORIES FOR THE PRODUCTION OF AROMATIC COMPOUNDS: FROM BUILDING BLOCKS TO PLANT SECONDARY METABOLITES

A review manuscript published in the Journal of Industrial Microbiology and Biotechnology

Authors: Miguel Suástegui[1] and Zengyi Shao [1]

[1]Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

Abstract

The aromatic amino acid biosynthesis pathway is a source to a plethora of commercially relevant chemicals with very diverse industrial applications. Tremendous efforts in microbial engineering have led to the production of compounds ranging from small aromatic molecular building blocks all the way to intricate plant secondary metabolites. Particularly, the yeast *Saccharomyces cerevisiae* has been a great model organism given its superior capability to heterologously express long metabolic pathways especially the ones containing cytochrome P450 enzymes. This review contains a collection of *state-of-the-art* metabolic engineering work devoted towards unraveling the mechanisms for enhancing the flux of carbon into the aromatic pathway. Some of the molecules discussed include the polymer precursor muconic acid, as well as important nutraceuticals (flavonoids and stilbenoids), and opium-derived drugs (benzylisoquinoline alkaloids).

1. Introduction

During the past two decades, the biotechnological production of petroleum-derived compounds has attracted worldwide attention, leading to extraordinary research for the construction and optimization of microbial factories capable of utilizing renewable feedstocks as starting materials (Becker and Wittmann, 2015; Borodina and Nielsen, 2014; Jullesson et al., 2015). Given the richness and variety of biological pathways, bio-based production by engineered microorganisms has expanded from just biofuels to the realm of specialty chemicals and secondary metabolites (Dai et al., 2014; Facchini et al., 2012). This group of chemicals has garnered special attention due to the cumbersome processes required for their production through classical chemical synthesis or the immense amounts of plant tissues required for extraction. Amongst the core metabolic pathways that lead to the production of specialty compounds, the aromatic amino acid biosynthetic pathway outstands for its unique diversity.

The aromatic, or arene, chemicals are recognized as cyclic six-carbon structures with delocalized electrons that generally possess a sweet or pleasant *aroma*, the property from which this class of molecules gets its name. This vast family of compounds is of great importance to the chemical industry; benzene, for instance, serves as the aromatic building block of many other chemicals such as toluene, phenol, and other polycyclic aromatics. Unfortunately, using petroleum-derived benzene as the starting molecule compromises the sustainability and the safety of the process. Acute and chronic exposure to benzene has proven to be carcinogenic (Sonawane et al., 2000); furthermore, the synthesis of benzene derivatives usually entails the release of toxic gases into the environment (Niu et al., 2002b). For these

reasons, the production of aromatic compounds in microbial systems has become more attractive.

The family of aromatic compounds comprises a vast number of molecules. In nature, the aromatic amino acids, namely, L-tryptophan (L-trp), L-phenylalanine (L-phe), and L-tyrosine (L-tyr), serve as building blocks for the biosynthesis of polypeptides. In several plants, these amino acids serve also as the precursors for the production of highly sought-after nutraceuticals, fragrances, and drugs. Although these secondary metabolites encompass a market in the billion-dollar range (Pandal, 2014), extraction from plant tissues requires high amounts of biomass and cumbersome separation processes, leading to very low yields. Hence, the implementation of microbial factories, engineered with the heterologous biosynthetic pathways of aromatic secondary metabolites, represents a great solution to this problem. This ensures not only higher yields but also sustainable and greener processes. With the advent of recombinant DNA technology, genome editing tools, and synthetic biology, this solution has proven to be feasible (Becker and Wittmann, 2015; Borodina and Nielsen, 2014; Jullesson et al., 2015), albeit, optimization of the microbial hosts to increase the carbon yields still remains as a limiting factor.

This review summarizes the metabolic engineering efforts to manipulate the biosynthesis of aromatic compounds in microbial systems with exclusive attention directed to *Saccharomyces cerevisiae*. Yeast hosts have a greater potential in the production of secondary aromatic compounds due to their greater capability of expressing membrane-bound cytochrome P450 oxidases, which are key catalysts in these metabolic pathways (Jiang and Morgan, 2004; Schoch et al., 2001; Szczebara et al., 2003).

2. The Aromatic Amino Acid Biosynthetic Pathway

The aromatic amino acid biosynthetic pathway is present in many microorganisms and plants; it is responsible for the *de novo* synthesis of L-tyr, L-phe, and L-trp. This pathway is sectioned into three main parts: (i) the shikimic acid pathway, (ii) the L-trp branch, and (iii) the L-tyr and L-phe branches (**Figure 1**). The energetic expenditure of synthetizing aromatic acids requires twice the ATP cost than most of other amino acids (Bender, 2012). Therefore, it is expected to encounter strict regulation at the transcriptional and allosteric levels throughout the pathway.

2.1. Initiation, shikimic acid pathway, and aromatic amino acid branches

The initiation of the synthesis of aromatic amino acids requires two precursors from distinct, yet closely related pathways. Erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) are the two starter units; the former derives from the pentose phosphate pathway (PPP) and the latter from glycolysis. A balanced ratio between these two molecules would be optimal to enhance the carbon entrance into the aromatic pathway. However, given the intrinsic functional differences between these parallel pathways, the available fluxes of these individual precursors differ drastically. Metabolic flux analysis in *S. cerevisiae* has shown that the available E4P flux is at least an order of magnitude lower than the PEP flux (Suastegui et al., 2016a). Hence, as described in the subsequent sections, efforts have been directed towards engineering of PPP as well as the lower end of the glycolytic pathway to enhance the production of precursors to enter the aromatic pathway.

The enzyme 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthase (isoenzymes *ARO3/4)* is in charge of condensing the two precursors to initiate the biosynthesis of aromatic compounds. Following this reaction, the pentafunctional *ARO1* enzyme catalyzes five out of

the six steps in the shikimic acid pathway, from DAHP to 5-enolpyruvyl-3-shikimate phosphate (EPSP). This enzyme is exemplary in yeast since all five subunits are encoded within one open reading frame (ORF), resulting in single multi-domain protein, unlike its homologues in bacteria and plants, where each subunit is encoded separately. Finally, the enzyme chorismate synthase *ARO2* converts EPSP into chorismate, which is the common precursor to the biosynthesis of L-trp, L-phe, and L-tyr (Braus, 1991). Chorismate can enter the L-trp branch composed of five steps (*TRP1* to *TRP5)* or enter the L-tyr/L-phe branches which share the first catalytic step, from chorismate to prephenate, catalyzed by the enzyme chorismate mutase *ARO7*. Following this step, the pathway branches into the synthesis of phenylpyruvate (PP), catalyzed by prephenate dehydratase (*PHA2*); and the synthesis of 4-hydroxyphenylpyruvate, (4-HPP) catalyzed by prephenate dehydrogenase (*TYR1*). Finally, two aromatic amino transferases, *ARO8* and *ARO9,* catalyze the transfer of an amine group to each of the previous molecules to yield L-phe and L-tyr (Braus, 1991) (**Figure 1**).

2.2. Transcriptional regulation

It is well known that transcription factors (TFs) can globally regulate the transcription rates of a network of genes. TFs bind to specific promoter regions of structural genes to up- or down-regulate the transcription of such genes (Zhu et al., 2009). GCN4p is a TF that has been associated with transcriptional activation of amino acid biosynthesis genes in response to nutrient starvation (Hinnebusch, 1988; Hinnebusch and Natarajan, 2002). This regulatory protein contains a leucine-zipper structure that is commonly found in DNA-binding polypeptides. As expected, several GCN4p binding sequences have been found in the promoter sequences of the aromatic amino acid biosynthesis genes such as *ARO1*, *ARO3, ARO4, ARO9, TRP2,* and *TRP3* (Braus, 1991). It has been demonstrated that certain metabolites can control

the transcription rates of specific genes in the pathway. For instance, the transcription of *TYR1* can be regulated by the concentrations of L-phe (Mannhaupt et al., 1989). Another example is the influence of excess L-trp in the activation of *ARO9* and *ARO10,* which can initiate the process of aromatic amino acid degradation (Lee and Hahn, 2013). The implementation of a global transcriptional regulation approach to increase the production of aromatic compounds, although promising, is naturally challenging. For instance, the starvation of aromatic amino acids will not only trigger a GCN4p-dependent transcriptional increase response on the genes from the corresponding pathway, but other genes involved in the biosynthesis of histidine, arginine, isoleucine, leucine, and other amino acids will also be activated, preventing the possibility of conducting a controllable localized up-regulation (Hinnebusch and Natarajan, 2002). This creates intricate regulatory networks that have to be explored from a global perspective (Moxley et al., 2009).

2.3. Allosteric regulation

There are three major allosteric checkpoints occurring at the branching points of the aromatic pathway. Engineering these control points in *S. cerevisiae* is the most straightforward approach and thus has enabled the heterologous production of aromatic compounds from the three main tiers in the pathway. The first one occurs at the level of the DAHP synthase isoenzymes *ARO3* and *ARO4* (Luttik et al., 2008b). These catalyze the first committed step in the biosynthesis of aromatic amino acids, and both *ARO3* and *ARO4* are sensitive to the feedback inhibition posed by L-phe or L-tyr, respectively (**Figure 1**). Thus, the presence of these two amino acids can block the entrance of carbon into the pathway, dramatically decreasing the production of any of the downstream metabolites.

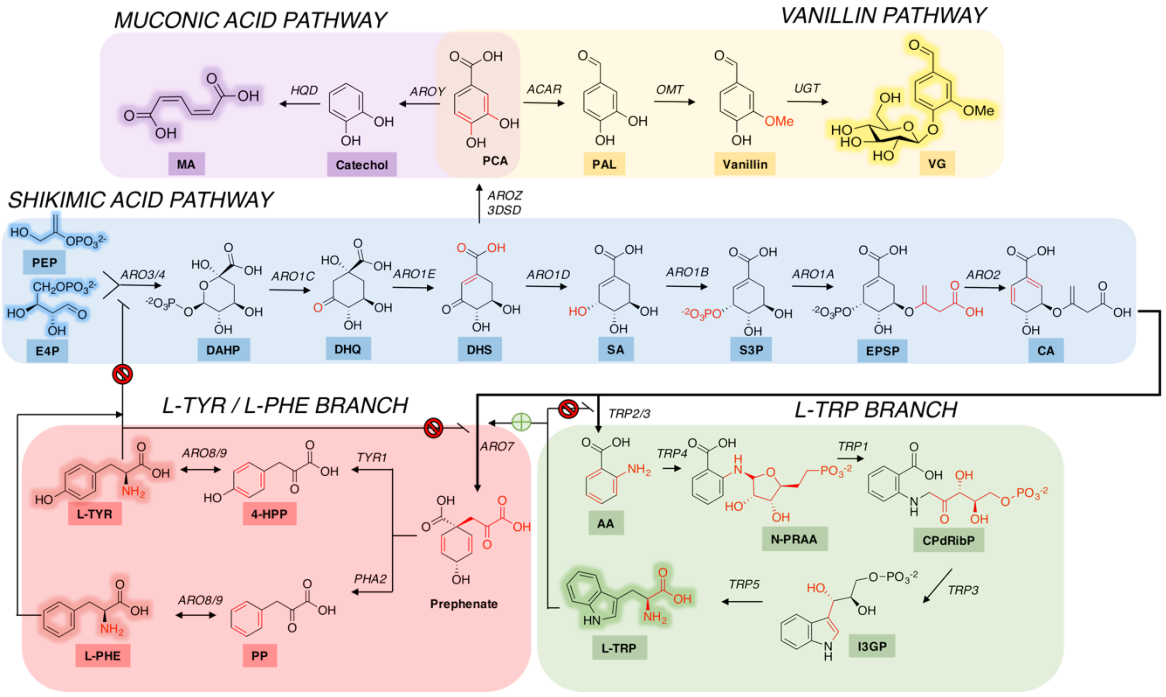


Figure 1: The aromatic amino acid biosynthetic pathway, the heterologous muconic acid (MA) pathway, and the heterologous vanillin pathway. Abbreviation of metabolites – E4P, erythrose-4-phosphate; PEP, phosphoenolpyruvate; DAHP, 3-deoxy-D-arabinoheptulosonate-7-phosphate; DHQ, dehydroquinoate; DHS, dehydroshikimate; SA, shikimic acid; S3P, shikimate-3-phosphate; EPSP, 5-enolpyruvyl-3-shikimate phosphate; CA, chorismic acid; PP, phenylpyruvate; 4-HPP, *p*-hydroxyphenylpyruvate; L-PHE, L-phenylalanine; L-TYR, L-tyrosine; AA, anthranilic acid, N-PRAA, N-(5-phospho-D-ribosyl)-anthranilate; CPdRibP, 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate; I3GP, indole-3-glycerol phosphate; L-TRP, L-tryptophan; PCA, protocatechuic acid; PAL, protocatechuic aldehyde; VG, vanillin-β-D-glucoside. Abbreviation of enzymes – *ARO3/4,* DAHP synthase isoenzymes; *ARO1*, aromatic pentafunctional enzyme; subunits – *C*: DHQ synthase, *E*: DHQ dehydratase, *D*: DHS dehydrogenase, *B*: SA kinase, *A*: ESPS synthase; *ARO2*, chorismate synthase; *ARO7*, chorismate mutase; *PHA2,* prephenate dehydratase; *TYR1,* prephenate dehydrogenase; *ARO8/9;* aromatic amino transferases; *TRP2*, anthranilate synthase; *TRP3*, indole-3-glycerol-phosphate synthase; *TRP4*, anthranilate phosphoribosyl transferase; *TRP1*, phosphoribosylanthranilate isomerase; *TRP5*, tryptophan synthase; *AROZ* and 3DSD, 3DHS dehydratase; *AROY*, PCA decarboxylase; *HQD*, catechol 1,2-dioxygenase; *ACAR*, aromatic carboxylic acid reductase; *OMT*, *O*-methyl transferase; *UGT*, UDP glycotransferase.

The second allosteric control occurs on the enzymes anthranilate synthase (*TRP2*) and indole-3-glycerol-phosphate synthase (*TRP3*), whose activities are repressed by the presence of L-trp. Simultaneously, L-trp can act as an activator of the enzyme chorismate mutase (*ARO7*), which catalyzes the conversion of chorismate into prephenate, initiating the production of L-tyr and L-phe (**Figure 1**). A third recognized allosteric control occurs at high concentrations of L-tyr. This causes feedback inhibition of *ARO7*, potentially switching the carbon flux towards production of L-trp to maintain a balanced production of aromatic amino acids (Braus, 1991).

3. Metabolic Engineering Strategies for the Production of Aromatic Compounds

3.1. Engineering the shikimic acid pathway

The shikimic acid pathway has been a focal point in *S. cerevisiae* to engineer the production of three important metabolites: muconic acid, shikimic acid, and vanillin.

3.1.1. Muconic acid

Muconic acid is the unsaturated precursor of adipic acid and terephthalic acid, which are the building units for the plastics Nylon 6,6 and polyethylene terephthalate (PET) with an annual market value greater than $22 million (Lu et al., 2016; Niu et al., 2002b; Picataggio and Beardslee, 2012). The classic chemical processes to produce these derivatives are not environmentally friendly, thus stressing the need for development of microbial fermentation alternatives. Curran *et al.* (Curran et al., 2012a) engineered the first *S. cerevisiae* strain capable of accumulating this compound by introducing a 3-gene heterologous pathway which diverted carbon from the shikimate pathway at the node of dehydroshikimate (DHS) (**Figure 1**). The

metabolic engineering rationale included the removal of the feedback inhibition loop by overexpressing the L-tyr insensitive DAHP synthase variant, $ARO4_{K229L}$ (**Figure 2a**). Furthermore, to increase the precursor availability, flux balance analysis (FBA) was performed. The results suggested that the PPP had to be rewired to force the carbon entrance through the non-oxidative PPP branch, allowing a higher E4P pool to enter the aromatic pathway. The overexpression of the transketolase gene (*TKL1*) and the deletion of glucose-6-phosphate-1-dehydrogenase (*ZWF1*) were the key predicted manipulations to achieve this goal (**Figure 2b**). These manipulations translated into an increase in muconic acid titer from 40 mg L^{-1} to ~60 mg L^{-1} (Curran et al., 2012a). Overexpression of the rate-limiting step, protocatechuic acid (PCA) decarboxylase (*AROY*), led to a final strain producing 141 ± 8 mg L^{-1} (3.9 mg $g^{-1}_{glucose}$) (**Table 1**). A second study demonstrated that overexpressing the mutant $ARO1_{D1409A}$, capable of halting the conversion of DHS to shikimic acid, increased the titer to 235 mg L^{-1} (**Figure 2c**). Coupling this strategy with a controlled oxygen fermentation to relieve the oxygen sensitivity of the *AROY* isolated from the anaerobic bacterium *Klebsiella pneumoniae* increased the muconic acid titer to 559.5 mg L^{-1}, the highest reported titer and yield to date in *S. cerevisiae* (Suastegui et al., 2016b) (**Table 1**).

3.1.2. Shikimic acid

Shikimic acid is the fourth intermediate in the shikimate pathway, and serves as the precursor of the commercial anti-influenza drug Tamiflu® (Adelfo Escalante et al., 2014). It is commonly extracted from plants (the genus *Illicium*), however, the low yield, long cultivation cycle, and susceptibility of open-field crop growth to environmental factors pose the microbial production as a preferred alternative.

The engineering key to enable the accumulation of shikimic acid in *S. cerevisiae* is the overexpression of the mutant enzyme $ARO1_{D920A}$, which has the shikimate kinase subunit disrupted and hence cannot convert shikimic acid into shikimate-3-phosphate (S3P) (**Figure 1 and 2c**). A recent study compared the capacity of four distinct *S. cerevisiae* strains to accumulate SA as an approach to model the carbon entrance into the aromatic amino acid pathway (**Figure 2d**). It was demonstrated that the four equally engineered strains, containing the same set of episomal plasmids for gene overexpression, differed dramatically in their titers.

Table 1: *De novo* production of the most significant aromatic amino acids-derived compounds in *S. cerevisiae.*

	Compound	Strain	Substrate	Titer mg L^{-1}	Yield mg g^{-1}	Fermentation Type	Ref
Shikimic acid pathway	Shikimic acid	INVSc1-SA3	Glucose	380	19	Batch	(Suastegui et al., 2016a)
	Muconic acid	MA-12	Glucose	141	3.9	Batch	(Curran et al., 2012a)
		INVSc1-MA2	Glucose	559.5	14	Batch (mini reactor with 10-20% dissolved oxygen)	(Suastegui et al., 2016b)
	Vanillin	VG2	Glucose	500	32	Continuous chemostat (dilution rate = 0.015 h^{-1})	(Brochado et al., 2010)
L-tyr / L-phe branch	L-tyrosine	TY1041M	Glucose	350 (cytosolic)	17	Batch	(Gold et al., 2015)
	Coumaric acid	ST4058	Glucose	1930	N.S.	Fed batch (FIT media)	(Rodriguez et al., 2015c)
	Naringenin	IMX106	Glucose	108	5.4	Batch	(Koopman et al., 2012b)
	trans-Resveratrol	ST4152	Glucose	415.6	N.S.	Fed-batch	(Li et al., 2015)
			Ethanol	531.4	N.S.		
	Thebaine	CSY1064	Glucose	0.0064	0.00032	Batch	(Galanie et al., 2015)
	Hydrocodone	Pcs2765	Glucose	0.0003	0.000015	Batch	
L-trp branch	N-Acetylserotonin	SCE-Il3-HM-60	Glucose	~ 45	N.S.	Fed-batch (FIT media)	(Germann et al., 2015)
	Melatonin	SCE-Il3-HM-60	Glucose	14.5	N.S.	Fed-batch (FIT media)	

N.S.: Not specified

Metabolic flux analysis (MFA) coupled with quantitative PCR (Qpcr) enabled the establishment of an appropriate engineering rationale that allowed titer increases in all four strains. This strategy consisted of grouping the genes encoding *TKL1,* $ARO4_{K229L}$, and $ARO1_{D920A}$ together in a low-copy plasmid, demonstrating that the aromatic amino acid pathway is sensitive to the burden of excessive protein expression (Suastegui et al., 2016a). The highest shikimic acid titer in batch fermentation reached 380 mg L^{-1} in the diploid strain INVSc1 (**Table 1**).

3.1.3. Vanillin

The production of aromatic flavors or aromas is predominant in several industries. The two most popular benzenoid flavors are vanillin and 2-phenylethanol (derived from L-phe, section 3.3.1); the market of these two compounds together make 44.4% of the total benzenoid market, which as to 2014 was $2.8 billion, and is expected to increment to $3.4 billion by 2019 (Pandal, 2014). Vanillin is the most extensively used flavoring compound in industry. Its global market in 2019 is expected to reach $800 million, which represents the highest increase among all the benzenoid flavors. This plant-derived secondary metabolite can be extracted from the seedpods of the flower *Vanilla planifolia*. However, due to extremely low yields and laborious cultivation schemes that lead to a cost-ineffective process, nearly 99% of vanillin is obtained from a biochemical synthesis processes involving the breakdown of lignocellulosic biomass or from a petroleum derived precursor (Priefert et al., 2001).

Microbial production of vanillin from glucose has also been reported in *S. cerevisiae via* expression of a heterologous pathway that diverts the carbon from the shikimic acid pathway at the node of DHS. This pathway consisted of four steps, namely, DHS dehydratase (*3DSD*) from *Podospora pausiceta*, an aromatic carboxylic acid reductase (*ACAR*) from *Nocardia sp.*,

a phosphopantetheinyl transferase (*PPTase*) from *Escherichia coli* to activate *ACAR*, and an *O*-methyl transferase (*OMT*) from *Homo sapiens* (**Figure 1**). The introduction of this pathway in *S. cerevisiae* resulted in the accumulation of 45 mg L^{-1} of vanillin after 48 h of small-scale batch fermentation (Hansen et al., 2009b). A further study addressed the toxicity of vanillin by overexpression of a glycotransferase (*UGT*) from *Arabidopsis thaliana* in the previously reported strain, yielding a strain capable of producing 100 mg L^{-1} of vanillin β-D-glucoside (VG) after 90 h of fermentation (Brochado et al., 2010). Furthermore, the same group investigated engineering strategies to increase the carbon flux into the aromatic pathway by means of *in silico* metabolic modeling. The FBA prediction suggested that the knockout of *GHD1* encoding glutamate dehydrogenase could enhance the production of VG. This was supported by the fact that deletion of this gene increased the pool of NADPH, thermodynamically favoring the activity of *ACAR*.

However, this prediction did not result in higher vanillin titers due to a reduced biomass yield on glucose. A second knockout target was the gene *PDC1* encoding pyruvate decarboxylase, whose deletion could reduce the overall pyruvate decarboxylase activity by ~30%, ensuring that less carbon was directed towards production of ethanol while maintaining cellular growth (**Figure 2b**). This resulted in a 2-fold increased production of VG compared to the parental strain from 250 to 500 mg L^{-1} in a very low dilution rate chemostat fermentation (Brochado et al., 2010) (**Table 1**).

3.2. Engineering the L-tyr and L-phe branches

In plants, the L-tyr and L-phe metabolic branches provide precursors to a great amount of aromatic secondary metabolites such as flavonoids, stilbenoids, and alkaloids. The application of these molecules ranges from the food and nutraceutical industries to the pharmaceutical industry. For instance, the pain management pharmaceuticals, a group of drugs that contain the L-tyr-derived alkaloids, are expected to reach a global market of $40.8 billion by 2020 (Elder, 2015). Extraction of these compounds from plant tissues, however, is strongly dependent on environmental factors that can cause production instability in the farming processes. Hence, the production of secondary metabolites derived from L-tyr and L-phe has been intensively studied in microbial platforms in the past decade.

To enable its microbial production, research has been directed mostly towards engineering the aromatic amino acid pathway for accumulation of L-tyr. The work by Gold *et al*. (Gold et al., 2015) described a metabolomics approach aided by *in silico* analysis for the construction of a yeast platform for overproduction of L-tyr. To construct the base strain, the feedback inhibition points were removed by overexpression of the *S. cerevisiae* feedback insensitive enzymes $ARO4_{K229L}$ and $ARO7_{G141S}$ (**Figure 2a**), and the byproduct formation was removed by knocking out phenylpyruvate decarboxylase *ARO10* (**Figure 2c**). Especial attention was directed towards engineering the central carbon metabolism to enhance the production of precursors. Deletion of *ZWF1* encoding glucose-6-phosphate dehydrogenase was attempted with the hypothesis of linking NADPH deficiency and subsequent regeneration in the L-tyr pathway by the action of overexpressed $NADP^+$-dependent prephenate dehydrogenase *TYR1*. The L-tyr concentrations were improved to 192 Mm (~350 mg L^{-1}) when L-methionine was added into the medium to recover the growth deficiency caused by deletion of *ZWF*1 because

L-methionine reduces the cellular demand for NADPH during amino acid synthesis (**Table 1**). Finally, to increase availability of PEP, a mutant version of the major pyruvate kinase (*CDC19*$_{T21E}$) (**Figure 2b**), with impaired activity was knocked in to substitute the wildtype gene. However, no increase of carbon flux into the aromatic amino acid pathway was observed. Even though the mutant still possessed low pyruvate kinase activity, the decreased cellular fitness made the cells unviable in the absence of L-methionine and caused the L-tyr production to drop significantly.

3.2.1. 2-Phenylethanol

Phenethyl alcohol (2-PE) is a benzene-derived chemical with an ethylic group widely used in industry due to its particular floral odor. Altogether, its market is expected to reach $700 million by 2019 (Pandal, 2014), yet its current production process is based on chemical synthesis (alkylation of benzene) which yields undesired side products, making the separation processes costly (Etschmann et al., 2002). Production of 2-PE has also been studied in microbial systems. As a naturally occurring molecule in *S. cerevisiae,* it provides organoleptic features to fermented beverages like sake and beer. This compound can be produced *de novo* through the conversion of glucose to L-phe. The latter amino acid undergoes transamination, decarboxylation, and dehydrogenation to finally yield 2-PE (a mechanism known as the Ehrlich pathway) (**Figure 3**). Work has been directed towards increasing the *de novo* production of 2-PE. For instance, Fukuda *et al.* (Fukuda et al., 1992) subjected *S. cerevisiae* to inhibitory concentrations of fluorophenylalanines, which act as competitive inhibitors of L-phe synthesis. The recovered spontaneous mutants, those with higher tolerance to the inhibitors could also accumulate 6-fold higher 2-PE (1.3 g/L) than the parental strain (Fukuda et al., 1990b). Analysis of the mutant strain revealed a mutation on the L-tyr-inhibited DAHP

synthase (*ARO4*) causing its release from the feedback inhibition and resulting in higher carbon flux towards production of L-tyr, L-phe and 2-PE (Fukuda et al., 1990a).

Biotransformation of L-phe to produce 2-PE has also been extensively studied in *S. cerevisiae* as an alternative to overcome the low yields of *de novo* synthesis. This process requires the addition of L-phe into the fermentation medium, usually accompanied by a poor nitrogen source, forcing the cells to catabolize the aromatic amino acid. Concentrations as high as 4.8 g L^{-1} have been obtained in yeast (Kim et al., 2014a), albeit product toxicity remained as the limiting factor. *In situ* product recovery has thus been crucial in the production of 2-PE, with the most successful technique being two-phase extraction using oleic acid as the extractant, which elevated the production to 12.6 g L^{-1} (Stark et al., 2003; Stark et al., 2002).

3.2.2. Coumaric acid

The production of the important nutraceutical group of flavonoids and stilbenoids initiates with the common precursor *p*-coumaric acid (Phca). This compound can be derived directly from L-tyr through the enzyme L-tyr ammonia lyase (*TAL*), or from L-phe, through the sequential action of two enzymes: phenylalanine ammonia lyase (*PAL*) and cinammate-4-hydroxylase (*C4H*) (Vannelli et al., 2007) (**Figure 3**).

Recently, Jendresen *et al*. (Jendresen et al., 2015) characterized several *TAL*s from various microbial sources, resulting in the finding of two enzymes, from the bacteria *Herpetosiphon aurantiacus* and *Flavobacterium johnsoniae,* with increased catalytic activity and high product specificity (barely any cinnamic acid was observed). The latter enzyme (*FjTAL*) was incorporated in the subsequent research by Rodriguez *et al.* (Rodriguez et al., 2015c) with the goal of constructing a *S. cerevisiae* platform for high accumulation of Phca. Unlike the work described by Gold *et al.* (Gold et al., 2015), the focus was only directed to manipulate the

aromatic amino acid pathway without altering precursor supply through tailoring the central carbon metabolism

Similar strategies were implemented: removal of L-tyr feedback inhibition (overexpression of $ARO4_{K229L}$ and $ARO7_{G141S}$), removal of competitive pathways leading to phenylacetaldehyde (PAA) and 4-hydroxyphenylacetaldehyde (4-HPAA) (*ARO10* and pyruvate decarboxylase *PDC5*; **Figure 2c**), and channeling of carbon flux towards L-tyr (overexpression of *TYR1*, **Figure 1**). Finally, to study the push of carbon flux through the pathway, the complete shikimate pathway from *S. cerevisiae* (*ARO1* and *ARO2*) and the orthologous genes from *E. coli* were independently integrated into the genome under the control of strong constitutive promoters. Overexpression of the *E. coli* SA kinase II alone (*aroL*) demonstrated the biggest improvement, indicating that this step could be rate-limiting in the pathway. The highest producing strain accumulated 1.93 g L^{-1} of Phca in microplate fermentation and 1.89 g L^{-1} in bioreactors, both using synthetic Feed-In-Time (FIT) fed-batch media that gradually release glucose from higher polysaccharides within the medium solution under the action of supplemented hydrolase (**Table 1**).

.

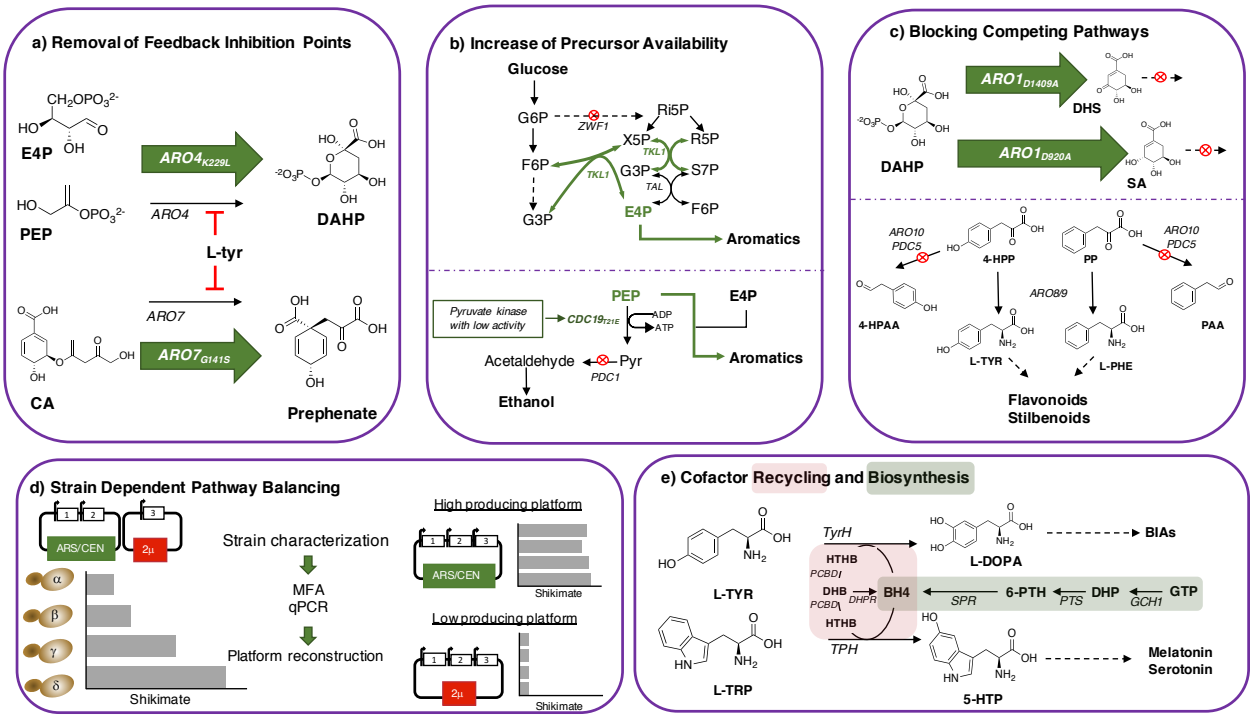


Figure 2: **Common metabolic engineering strategies implemented in *S. cerevisiae* to increase the production of compounds deriving from the aromatic amino acid biosynthetic pathway**. A) Removal of feedback inhibition points has been achieved by overexpression of the tyrosine insensitive enzymes $ARO4_{K229L}$, which deregulates the first committed step in the aromatic biosynthesis pathway; and $ARO7_{G141S}$, which controls the lower branching point to the production of the L-phe and L-tyr. B) Increasing the availability of the main precursor E4P has been achieved by overexpression of the transketolase 1 gene (*TKL1*) in combination with deleting the glucose-6-phosphate-1-dehydrogenase gene (*ZWF1*) to force the entrance of carbon into the non-oxidative section of PPP. Reducing the activity of the catalytic steps in the lower portion of glycolysis (pyruvate kinase mutant $CDC19_{T21E}$ and *PDC1*) has been attempted to increase the availability of PEP. C) Blocking competing pathways has been demonstrated by overexpressing mutant versions of the *ARO1* enzyme to halt the conversion of intermediates in the shikimate pathway. Deletion of *ARO10* and *PDC5* increases precursor availability. D) A strain-dependent pathway balancing approach was implemented to establish a strong yeast platform for the production of the common precursor shikimic acid. E) Recycling of the cofactor BH4 (tetrahydrobiopterin) is essential in the production of secondary aromatic compounds. Abbreviations – *DHB*, dihydrobiopterin; *6-PTH*, 6-pyruvoyl-tetrahydropterin; *DHP*, dihydroneopterin triphosphate; *GTP*, guanosine triphosphate; *PCBD*, pterin-4-alpha-carbinolamine dehydratase; *DHPR*, dihydropteridine reductase; *SPR*, sepiapterin reductase; *PTS*, 6-pyruvoyl-tetrahydropterin synthase; *GCH1*, GTP cyclohydrolase I. The dash lines represent multiple steps.

3.2.3. Flavonoids and stilbenoids

Jiang *et al.* reported the production of the flavanone nutraceutical naringenin in *S. cerevisiae* (Jiang et al., 2005). This work was further extended by Trantas *et al.* (Trantas et al., 2009) *via* constructing a series of strains capable of converting L-phe to a palate of flavonoids and stilbenoids, namely, *trans*-resveratrol, naringenin, genistein, kaempferol, and quercetin. To convert L-phe into the aforementioned products, initially, a strain overexpressing *PAL, C4H* (cytochrome P450 enzyme) and *CPR* (cytochrome P450 reductase) was constructed to accumulate Phca. Although *S. cerevisiae* contains a copy of *CPR,* overexpression of a *Glycine max* (soybean) *CPR* increased the accumulation of Phca by 4-fold. All of the subsequent strains contained a 4-coumaric acid ligase gene (*4CL*) to convert Phca to 4-coumaroyl CoA. To enter the stilbenoids pathway, the type III polyketide synthase (resveratrol synthase, *RS*) was overexpressed to produce *trans*-resveratrol. To construct the flavonoid-producing strains, the chalcone synthase (*CHS*) and the chalcone isomerase (*CHI*) were overexpressed, leading to production of naringenin. Further strains diverting naringenin into genistein (isoflavone) or flavonols (kaempferol and quercetin) were obtained by expression of the isoflavone synthase (*IFS*), or the flavanone 3-hydroxylase (*F3H/F3'H*) together with the flavonol synthase (*FLS*) (**Figure 3**). Although low titers were obtained, all the products were observed in the medium when L-phe or any of the downstream metabolites were fed into the medium. This indicated that all the plant genes were functional in *S. cerevisiae* but that pathway optimization was required to remove bottlenecks and enable higher conversion efficiencies.

Metabolic engineering strategies for *de novo* production of naringenin were investigated by Koopman *et al.* (Koopman et al., 2012b). The common strategies for removal of feedback inhibition and byproduct formation were implemented in this work. Additionally, expressing

the corresponding ammonia lyase enzymes (*TAL* or *PAL*) to increase precursor availability enhanced the production of Phca. The bottleneck step downstream was partially removed by integrating into the genome additional copies of *CHS*, leading to a production of 134.5 μM naringenin (36 mg L^{-1}) in batch fermentation. Optimization of the fermentation process in a bioreactor led to a maximal production of 400 μM (108 mg L^{-1}) (**Table 1**).

Resveratrol is a stilbenoid that is highly sought-after due to its nutraceutical properties. The production in S. ce*revisiae* has been achieved by whole cell conversion of Phca (Becker et al., 2003; Trantas et al., 2009) or in complex media containing L-tyr (Shin et al., 2012; Wang et al., 2011). Although one of the highest-producing strains on record could accumulate up to 391 mg L^{-1} resveratrol, it also required the supplementation of Phca (Sydor et al., 2010). Recently, a *S. cerevisiae* strain was constructed for the *de novo* production of resveratrol in minimal media without the supplementation of L-tyr or L-phe (Li et al., 2015). This entailed the expression, under strong constitutive promoters, of the heterologous *TAL* gene (tyrosine ammonia lyase) from *Herpetosiphon aurantiacus*, the *ST1* gene (resveratrol synthase) from *Vitis vinifera,* the *4CL* from *Arabidopsis thaliana,* and the acetyl-CoA carboxylase variant $ACC1_{S659A,\ S1157A}$ (post-translational de-regulated mutant) from *S. cerevisiae* to increase the availability of malonyl-CoA. These manipulations, together with the removal of L-tyr feedback inhibition led to the production of 235 mg L^{-1} resveratrol. Furthermore, production in fed batch mode using glucose or ethanol as feeding substrates after initial carbon depletion led to the production of 415.65 mg L^{-1} and 531.41 mg L^{-1} resveratrol, respectively, representing the highest productions reported to date (**Table 1**).

3.2.4. Benzylisoquinoline alkaloids

Benzylisoquinoline alkaloids (BIAs) compose a large group of plant secondary metabolites that include the opioid chemicals. Used for the treatment of severe pain, this group of molecules, which also contains codeine, morphine, oxycodone, and hydrocodone, represents the largest global market of pain management pharmaceuticals (Elder, 2015). Other BIAs like berberine and sanguinarine can be implemented as antibacterial agents (Fossati et al., 2014). *Papaver somniferum*, commonly named opium poppy, is the primary source of natural and semisynthetic opioids. Industrial farming of this plant to satisfy its huge demand, however, is prone to commercial instabilities due to its dependence on environmental factors. Hence, efforts to decipher the natural enzymatic pathways and construct microbial factories to produce these important metabolites to ensure a more controllable production scheme have been a work in progress for the last couple of decades (Beaudoin and Facchini, 2014; Hagel and Facchini, 2013; Ziegler et al., 2009). Eukaryotic microorganisms like *S. cerevisiae* have been central in this research, given their outstanding ability to express challenging heterologous enzymes, especially cytochrome P450s, whose activities require the presence of an electron donor protein (*CPR*) as well as a compartmentalized cellular environment provided by endoplasmic reticulum membranes (Ro et al., 2002).

One of the biggest breakthroughs of 2015 was the demonstration of the complete biosynthesis of BIAs from glucose in yeast. This work, published by Galanie *et al.* (Galanie et al., 2015), is a product of over a decade of work from several research groups that intensively worked on reconstructing the entire pathway leading to the production of thebaine (an important precursor of morphinan opioids) and hydrocodone (an important semisynthetic opioid) from glucose. The biosynthesis of the initial intermediate (*S*)-reticuline was targeted

first, which started from the heterologous expression of a repression insensitive L-tyr hydroxylase variant ($TyrH_{R37E, R38E, W166Y}$) from *Rattus norvegicus* coupled with four enzymes from the same source for the biosynthesis of the electron carrier co-substrate tetrahydrobiopterin (BH4) and its recycling and salvage enzymes to synthesize L-DOPA (**Figure 3** and **Figure 2e**). Central carbon metabolism was further modified following the examples presented in the previous sections, namely, derepression of feedback inhibition by expressing L-tyr insensitive *ARO4* and *ARO7* ($ARO4_{Q166K}$ and $ARO7_{T226I}$), expression of *TKL1* and deletion of *ZWF1*. To convert L-DOPA to (*S*)-reticuline, seven heterologous genes and the endogenous *ARO10* for production of 4-HPAA were overexpressed. The genes were isolated from different plants and bacteria, including the L-DOPA decarboxylase (DODC), a norcoclaurin synthase (*NCS*), three methyltransferases characterized in previous works (Hawkins and Smolke, 2008; Minami et al., 2008a), and one cytochrome P450 enzyme (NMCH) associated with *CPR* (Trenchard et al., 2015).

Completion of the pathway to produce thebaine required the determination of the key enzymatic step responsible for the stereochemical conversion of (*S*)-reticuline to I-reticuline. This goal was achieved by three independent groups, which opened the gate to the complete biosynthesis of morphine (Farrow et al., 2015; Galanie et al., 2015; Winzer et al., 2015), leading to the discovery of 1,2-dehydroreticuline synthase-1,2-dehydroreticuline reductase fusion enzymes (DRS-DRR) (**Figure 3**). The conversion of I-reticuline to thebaine was followed by expression of an engineered salutaridine synthase (*SalSyn,* bottleneck step) in which the *N*-linked glycosylation was avoided by replacement of the *SalSyn* α-helices with those from cheilanthifoline synthase (*CFS*), which is homologous to *SalSyn* but will not be glycosylated when expressed in yeast (Galanie et al., 2015). Furthermore, codon-optimized

salutaridine reductase (*SalR*) and salutaridinol acetyltransferase (*SalAT*) (Higashi et al., 2010; Ziegler et al., 2009) were expressed to complete the pathway to thebaine. Finally, the expression of thebaine 6-*O*-demethylase (*T6ODM*) and morphine reductase (*morB*) led to the construction of a strain capable of accumulating hydrocodone at a titer of ~0.3 μg L^{-1} after 120 hours of fermentation directly from sugars (**Table 1, Figure 3**).

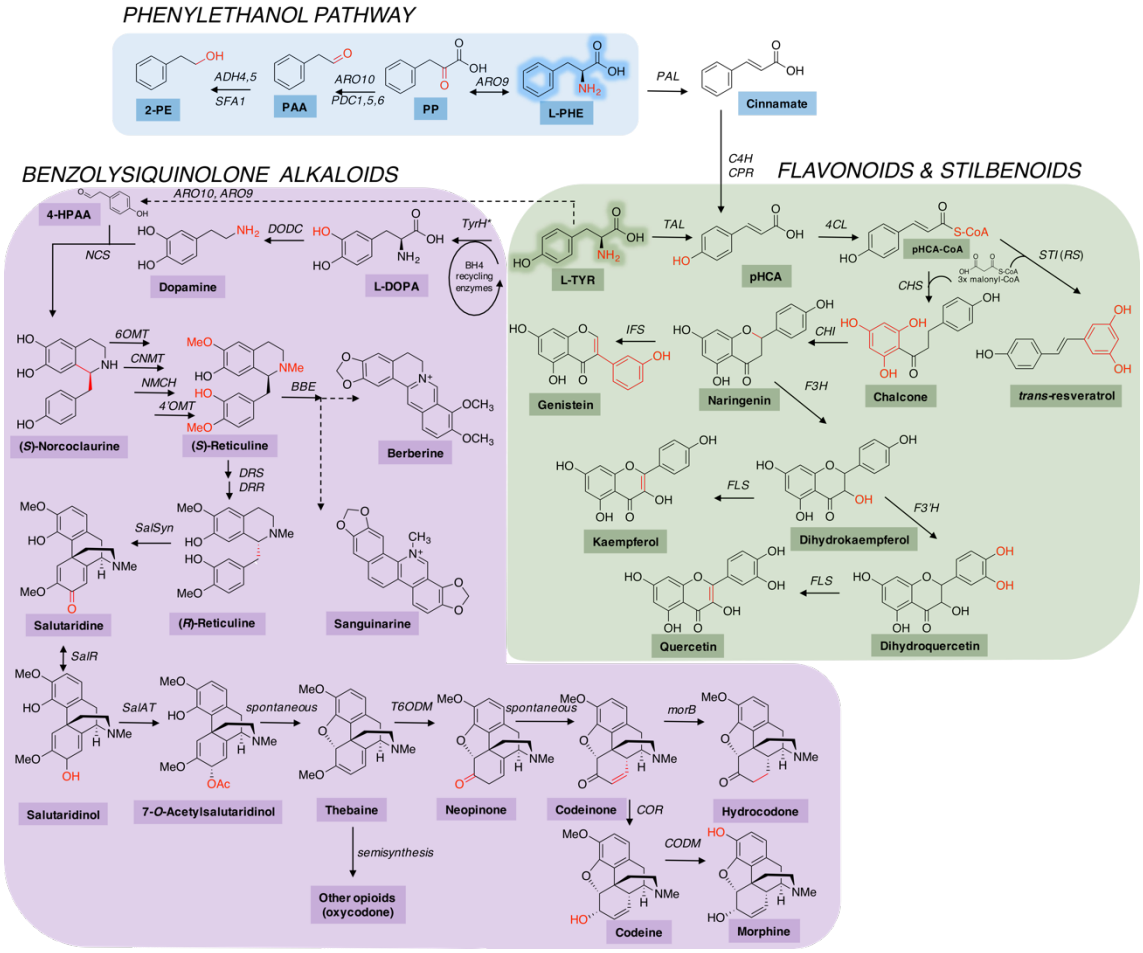


Figure 3: The metabolic pathways (native and heterologous) for production of L-tyr- or L-phe-derived aromatic compounds in *S cerevisiae*. Abbreviation of metabolites – PAA, phenylacetaldehyde; 2-PE, 30ccession3030ol; Phca, coumaric acid; L-DOPA, L-3,4-dihydroxyphenylalanine; 4-HPAA, 4-hydroxyphenylacetaldehyde. Abbreviation of enzymes – *ARO10*, phenylpyruvate decarboxylase; *PDC1,5,6,10*, pyruvate decarboxylase isoenzymes; *ADH4,5*, alcohol dehydrogenase isoenzymes; *SFA1*, bifunctional *ADH* and formaldehyde dehydrogenase; *PAL*, phenylalanine ammonia lyase; *TAL*, tyrosine ammonia lyase; *4CL*, 4-coumaric acid-CoA ligase; *RS*, resveratrol synthase; *CHS*, chalcone synthase; *CHI*, chalcone isomerase; *IFS*, isoflavone synthase; *F3H/F3'H*, flavanone 3-hydroxylase; *FLS,* flavonol synthase; *TyrH**, tyrosine hydroxylase repression insensitive mutant, *DODC*, L-DOPA decarboxylase; *morB*, morphine reductase. The following enzymes require cytochrome P450 reductase (*CPR*): *C4H,* cinammate-4-hydroxylase; *NCS*, norcoclaurine synthase; *6OMT*, norcoclaurine 6-*O*-methyltransferase; *CNMT*, coclaurine *N*-methyltransferase; *4'OMT*, 3'-hydroxy-*N*-methylcoclaurine 4'-*O*-methyltransferase; *NCMH*, *N*-methylcoclaurine hydroxylase; *BBE*, berberine bridge enzyme; *DRS*, 1,2-dehydroreticuline synthase; *DRR*, 1,2-dehydroreticuline reductase; *SalSyn*, salutaridine synthase; *SalR*, salutaridine reductase; *SalAT*, salutaridinol acetyltransferase; *T6ODM*, thebaine 6-*O*-demethylase; *COR*, codeinone reductase; *CODM*, codeine demethylase. The dash lines represent multiple steps.

3.3. Engineering the L-trp branch

The biosynthesis of L-trp initiates with the conversion of chorismate to anthranilate by action of the enzymatic complex *TRP2* and *TRP3*. The entrance of carbon into this branch of the aromatic amino acid pathway can be regulated by the concentrations of L-tyr and L-trp. L-tyr feedback inhibits *ARO7*, whereas L-trp counteracts this inhibition but feedback inhibits *TRP2* and *TRP3*, therefore allowing carbon flux to be appropriately split between two branches (Brown and Dawes, 1990) (**Figure 1**). Although L-trp is an important metabolite and key in the production of natural products (*e.g.* indole alkaloids) as well as mammalian neurotransmitters like melatonin and serotonin, so far there are no reports of a *S. cerevisiae* strain capable of overproducing L-trp.

3.3.1. Monoterpene indole alkaloids

Monoterpene indole alkaloids (MIAs) are a group of secondary metabolites found in many plants with diverse medicinal properties (Pasquali et al., 2006). Two common MIAs are the *Catharanthus roseus*-derived vinblastine and vincristine, which have been approved for anticancer treatment (Gobbi et al., 2003). Engineering the main MIA precursor strictosidine in yeast has been achieved by direct supplementation of the precursors tryptamine (a derivative of L-trp) and secologanin in a strain carrying the *C. roseus* strictosidine synthase, leading to a titer of 2 g L^{-1} strictosidine (Geerlings et al., 2001). *De novo* production of strictosidine has also been recently demonstrated. This work was devoted to engineering the mevalonate pathway for increased availability of geranyl pyrophosphate (GPP) followed by heterologous expression of 14 plant genes encoding the synthesis of secologanin. No evident engineering work was done on the aromatic amino acid pathway to increase the supply of L-trp for production of tryptamine (L-trp was supplemented in the amino acid mix of the growth

medium) (Brown et al., 2015). To date, engineering the production of L-trp has yet to be demonstrated.

3.3.2. Melatonin

Melatonin is a naturally occurring, L-trp-derived hormone in mammals that regulates the circadian cycle, and it is popularly known due to its nutraceutical properties as a sleeping aid as well as an antioxidant compound (Buscemi et al., 2004; Kamei et al., 2000). Although melatonin is commonly extracted from the pineal gland of mammalian organisms, it can also be naturally produced by *S. cerevisiae* during alcoholic fermentation. The production is highly dependent on the concentration of L-trp in the media, the different nitrogen sources, and the strain used (Rodriguez-Naranjo et al., 2012). As a GRAS (generally recognized as safe) microorganism, *de novo* production of melatonin in yeast could be preferred over extraction from animal tissues or from chemical synthesis which involved very toxic compounds (He et al., 2003).

Two groups have recently demonstrated the production of 5-hydroxytryptophan (5-HTP), which is the first intermediate in the synthesis of melatonin. Zhang *et al.* (Zhang et al., 2016) studied the action of three types of hydroxylases, *i.e.* L-trp hydroxylase (*T3H* or *T5H*), and L-phe hydroxylase *P4H*, on the conversion of L-trp to 5-HTP. Expression of *Oryctolagus cuniculus T5H* in combination with the biosynthesis pathway of the cofactor BH4 (**Figure 2e**) resulted in a strain producing around 7.3 $\mu g\ L^{-1}$ starting from 2 $g\ L^{-1}$ of L-trp.

A second group achieved *de novo* production of 5-HTP and also demonstrated the production of melatonin (Germann et al., 2015). The highest producing strain was constructed by integrating into the yeast's Ty2 retrotransposon sites multiple copies of L-trp hydrolysase from *Schistosoma mansoni* (*smTPH*), *H. sapiens* 5-HTP decarboxylase (*HsDDC*), *Bos 32ccess*

serotonin acetyltransferase (*BsAANAT*), and finally *H. sapiens* acetylserotonin *O*-methyltransferase (*HsASMT*) (**Figure 4**). BH4 was also recycled by expression of *P. aeruginosa* pterin-4-alpha-carbinolamine dehydratase (*PaPCBD1*), and *R. norvegicus* 6-pyruvoyl-tetrahydropterin synthase (*RnDHPR*) (**Figure 2e**). The highest producing strain accumulated melatonin at a titer of 14.5 mg L^{-1} in fed batch mode (**Table 1**). It is important to mention, however, that several strategies implemented in this work could not significantly elevate the production of melatonin; for instance, the external supplementation of L-trp did not show a significant increase. A potential approach suggested by the success in *E. coli* could be expression of the novel salicylate 5-hydroxylase, which can convert anthranilate into 5-hydroxyanthranilate. This molecule could further be processed through the promiscuous L-trp biosynthesis genes in *S. cerevisiae* leading to the direct production of 5-HTP, eliminating the usage of the unstable *T5H* and the requirement of expressing the enzymes for BH4 recycling (**Figure 2e**) (Sun et al., 2015). Given the minuscule titers observed from these previous examples, it is evident that engineering of the L-trp biosynthesis branch in *S. cerevisiae* is still in its infancy.

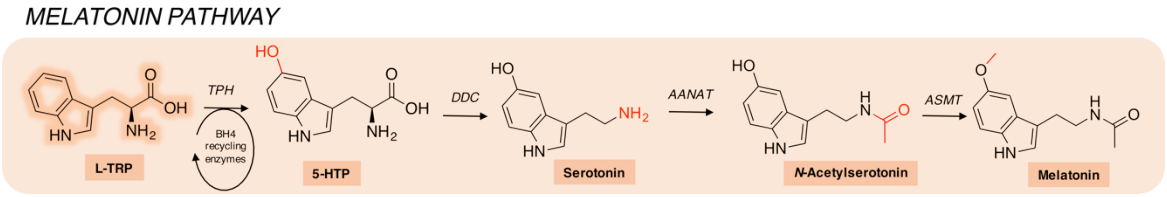


Figure 4: The heterologous metabolic pathway for production of L-trp-derived melatonin in *S. cerevisiae*. Abbreviation of metabolites – 5-HTP, 5-hydroxytryptophan. Abbreviation of enzymes – *TPH*, tryptophan hydroxylase; *DDC*, 5-hydroxy-L-tryptophan decarboxylase; *AANAT*, serotonin *N*-acetyltransferase; *ASMT*, *N*-acetylserotonin *O*-methyltransferase.

4. Remarks and Future Direction

The advent of synthetic biology, inexpensive DNA sequencing tools, and various bioinformatics tools has enabled the elucidation of complex aromatic amino acid-derived natural product pathways as well as the establishment of such pathways into microbial hosts for more efficient and sustainable production processes. However, it is evident that the yields and titers in *S. cerevisiae*, using a renewable sugar as the sole carbon source, are still far from meeting the market demands. As discussed in this review, the manipulations on the pathways to increase the carbon flow are merely limited to the removal of the feedback inhibition points (*ARO4* and *ARO7)* (**Figure 2**). Thus far, engineering availability of the precursor E4P has been modestly demonstrated by expression of *TKL1*, but altering the lower glycolytic pathway to enable higher accumulation of PEP for the production of DAHP remains as a great challenge. Furthermore, the examination of potential transcriptional regulators that could help rewire the carbon distribution in order to boost the carbon entrance into the aromatic pathway remains uninvestigated.

The most updated market analyses for aromatic compounds forecast considerable increases in national and worldwide demands in all major commercial segments, from food and nutraceuticals to drugs and pain management medications. This emphasizes the strong need for quick development of microbial hosts to gradually reduce the utilization of non-renewable resources as the main feedstocks. This goal can be accomplished by developing efficient (i) high-throughput screening methods, (ii) co-culture systems, and (iii) genetic tools for engineering non-conventional yeast.

The development of high-throughput screening methods allows exploration of large combinatorial spaces (strain libraries) for selection of improved phenotypes. Such methods can

be tied with a growth response (Du et al., 2012a) or an optical or electrochemical response (Bermejo et al., 2011; Heiskanen et al., 2013). One example of a biosensor to improve production of aromatic secondary metabolites was developed recently (DeLoache et al., 2015). It was based on screening a library of L-tyr hydroxylase mutants from *Beta vulgaris* (CYP74AD1) with enhanced activity. By coupling this enzyme with the L-DOPA dioxygenase from *Mirabilis jalapa*, L-DOPA was converted to a yellow fluorescent pigment betaxanthin. An identified mutant had a 2.8-fold improvement in activity, which translated into a 7.4-fold improvement in dopamine titer. This new mutant enzyme can be used to streamline the production of many valuable BIAs.

Co-culturing systems have garnered attention in recent years as an alternative to efficiently produce biorenewable chemicals (Zhang et al., 2015c). One advantage of such systems is the partitioning of large metabolic pathways into two or more microbial hosts, in order to reduce the metabolic burden imposed on a monoculture system, as well as to facilitate modular optimization. In the realm of aromatic compounds, Jones *et al*. (Jones et al., 2016) implemented an *E. coli* co-culturing process to improve the production of flavonoids. By dividing the pathway into two strains (malonyl-CoA requiring upstream and NADPH requiring downstream), and optimizing culturing conditions through computational modeling, they obtained a system with 65% improved titer in flavan-3-ols. Such an approach could also be implemented in yeast cultures with the advantage of a more efficient expression of cytochrome P450 enzymes.

Finally, although *S. cerevisiae* has been greatly domesticated and serves as a model for quick and robust engineering, a major disadvantage stems from being a Crabtree positive yeast. This entails the elevated production of ethanol, rather than biomass precursors, under aerobic

respiration in high glucose concentrations. As a result, lower titers for value-added compounds such as aromatic chemicals are observed. Although research has been devoted to diminish this effect in *S. cerevisiae* (Oud et al., 2012), a potential alternative solution is to engineer Crabtree negative non-conventional yeasts. Such yeasts include but are not limited to *Yarrowia lipolytica* and *Scheffersomyces stipitis.* Special attention should be directed to the latter species, since it possesses a highly efficient xylose utilization pathway. As a result, it is expected to carry a higher carbon flux into the PPP, which could ultimately lead to an enhanced production of aromatic compounds due to an elevated availability of the precursor E4P.

5. Bibliography

Adelfo Escalante, A., Carmona, S. B., Diaz Quiroz, D. C., Bolivar, F., 2014. Current perspectives on applications of shikimic and aminoshikimic acids in pharmaceutical chemistry. Res Rep Med Chem. 4**,** 35-46.

Beaudoin, G. A., Facchini, P. J., 2014. Benzylisoquinoline alkaloid biosynthesis in opium poppy. Planta. 240**,** 19-32.

Becker, J., Wittmann, C., 2015. Advanced biotechnology: metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. Angew Chem Int Ed Engl. 54**,** 3328-50.

Becker, J. V., Armstrong, G. O., van der Merwe, M. J., Lambrechts, M. G., Vivier, M. A., Pretorius, I. S., 2003. Metabolic engineering of *Saccharomyces cerevisiae* for the synthesis of the wine-related antioxidant resveratrol. FEMS Yeast Res. 4**,** 79-85.

Bender, D. A., 2012. Amino acid metabolism. Wiley-Blackwell, Chichester, West Sussex.

Bermejo, C., Haerizadeh, F., Takanaga, H., Chermak, D., Frommer, W. B., 2011. Optical sensors for measuring dynamic changes of cytosolic metabolite levels in yeast. Nat Protoc. 6**,** 1806-17.

Borodina, I., Nielsen, J., 2014. Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. Biotechnol J. 9**,** 609-20.

Braus, G. H., 1991. Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway. Microbiol Rev. 55**,** 349-70.

Brochado, A. R., Matos, C., Møller, B. L., Hansen, J., Mortensen, U. H., Patil, K. R., 2010. Improved vanillin production in baker's yeast through in silico design. Microb Cell Fact. 9**,** 84-84.

Brown, J. F., Dawes, I. W., 1990. Regulation of chorismate mutase in *Saccharomyces cerevisiae*. Mol Gen Genet. 220**,** 283-8.

Brown, S., Clastre, M., Courdavault, V., O'Connor, S. E., 2015. *De novo* production of the plant-derived alkaloid strictosidine in yeast. Proc Natl Acad Sci U S A. 112**,** 3205-10.

Buscemi, N., Vandermeer, B., Pandya, R., Hooton, N., Tjosvold, L., Hartling, L., Baker, G., Vohra, S., Klassen, T., 2004. Melatonin for treatment of sleep disorders. Evid Rep Technol Assess (Summ). 1-7.

Curran, K. A., Leavitt, J. M., Karim, A. S., Alper, H. S., 2012. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng. 15**,** 55-66.

Dai, Z., Liu, Y., Guo, J., Huang, L., Zhang, X., 2014. Yeast synthetic biology for high-value metabolites. FEMS Yeast Res. 15**,** 1-11.

DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J., Dueber, J. E., 2015. An enzyme-coupled biosensor enables (*S*)-reticuline production in yeast from glucose. Nat Chem Biol. 11**,** 465-71.

Du, J., Yuan, Y., Si, T., Lian, J., Zhao, H., 2012. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res. 40**,** e142.

Elder, M., The global market for pain management drugs and services. In: Research, B., (Ed.), Wellesley, MA, 2015.

Etschmann, M. M., Bluemke, W., Sell, D., Schrader, J., 2002. Biotechnological production of 2-phenylethanol. Appl Microbiol Biotechnol. 59**,** 1-8.

Facchini, P. J., Bohlmann, J., Covello, P. S., De Luca, V., Mahadevan, R., Page, J. E., Ro, D. K., Sensen, C. W., Storms, R., Martin, V. J., 2012. Synthetic biosystems for the production of high-value plant metabolites. Trends Biotechnol. 30**,** 127-31.

Farrow, S. C., Hagel, J. M., Beaudoin, G. A., Burns, D. C., Facchini, P. J., 2015. Stereochemical inversion of (*S*)-reticuline by a cytochrome P450 fusion in opium poppy. Nat Chem Biol. 11**,** 728-32.

Fossati, E., Ekins, A., Narcross, L., Zhu, Y., Falgueyret, J. P., Beaudoin, G. A., Facchini, P. J., Martin, V. J., 2014. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. Nat Commun. 5**,** 3283.

Fukuda, K., Watanabe, M., Asano, K., 1990a. Altered regulation of aromatic amino acid biosynthesis in β-phenylethyl alcohol overproducing mutants of sake yeast *Saccharomyces cerevisiae*. Agric Biol Chem. 54**,** 3151-3156.

Fukuda, K., Watanabe, M., Asano, K., Ouchi, K., Takasawa, S., 1992. Molecular breeding of a sake yeast with a mutated *ARO4* gene which causes both resistance to L-fluoro-DL-phenylalanine and increased production of β-phenethyl alcohol. Ferment and Bioeng. 73**,** 366-369.

Fukuda, K., Watanabe, M., Asano, K., Ueda, H., Ohta, S., 1990b. Breeding of brewing yeast producing a large amount of β-phenylethyl alcohol and β-phenylethyl acetate. Agric Biol Chem. 54**,** 269-271.

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., Smolke, C. D., 2015. Complete biosynthesis of opioids in yeast. Science. 349**,** 1095-100.

Geerlings, Redondo, Contin, Memelink, van der, H., Verpoorte, 2001. Biotransformation of tryptamine and secologanin into plant terpenoid indole alkaloids by transgenic yeast. Appl Microbiol Biotechnol. 56**,** 420-424.

Germann, S. M., Baallal Jacobsen, S. A., Schneider, K., Harrison, S. J., Jensen, N. B., Chen, X., Stahlhut, S. G., Borodina, I., Luo, H., Zhu, J., Maury, J., Forster, J., 2015. Glucose-based microbial production of the hormone melatonin in yeast *Saccharomyces cerevisiae*. Biotechnol J.

Gobbi, P. G., Broglia, C., Merli, F., Dell'Olio, M., Stelitano, C., Iannitto, E., Federico, M., Berte, R., Luisi, D., Molica, S., Cavalli, C., Dezza, L., Ascari, E., 2003. Vinblastine, bleomycin, and methotrexate chemotherapy plus irradiation for patients with early-stage, favorable Hodgkin lymphoma: the experience of the Gruppo Italiano Studio Linfomi. Cancer. 98**,** 2393-401.

Gold, N. D., Gowen, C. M., Lussier, F.-X., Cautha, S. C., Mahadevan, R., Martin, V. J. J., 2015. Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics. Microb Cell Fact. 14**,** 73.

Hagel, J. M., Facchini, P. J., 2013. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. Plant Cell Physiol. 54**,** 647-72.

Hansen, E. H., Moller, B. L., Kock, G. R., Bunner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S., Hansen, J., 2009. *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). Appl Environ Microbiol. 75**,** 2765-74.

Hawkins, K. M., Smolke, C. D., 2008. Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. Nat Chem Biol. 4**,** 564-73.

He, L., Li, J.-L., Zhang, J.-J., Su, P., Zheng, S.-L., 2003. Microwave Assisted Synthesis of Melatonin. Synth Comm. 33**,** 741-747.

Heiskanen, A., Coman, V., Kostesha, N., Sabourin, D., Haslett, N., Baronian, K., Gorton, L., Dufva, M., Emnéus, J., 2013. Bioelectrochemical probing of intracellular redox

processes in living yeast cells – application of redox polymer wiring in a microfluidic environment. Anal Bioanal Chem. 405**,** 3847-3858.

Higashi, Y., Smith, T. J., Jez, J. M., Kutchan, T. M., 2010. Crystallization and preliminary X-ray diffraction analysis of salutaridine reductase from the opium poppy *Papaver somniferum*. Acta Crystallogr Sect F Struct Biol Cryst Commun. 66**,** 163-6.

Hinnebusch, A. G., 1988. Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. Microbiol Rev. 52**,** 248-73.

Hinnebusch, A. G., Natarajan, K., 2002. Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. Eukaryotic Cell. 1**,** 22.

Jendresen, C. B., Stahlhut, S. G., Li, M., Gaspar, P., Siedler, S., Forster, J., Maury, J., Borodina, I., Nielsen, A. T., 2015. Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and *Saccharomyces cerevisiae*. Appl Environ Microbiol. 81**,** 4458-76.

Jiang, H., Morgan, J. A., 2004. Optimization of an in vivo plant P450 monooxygenase system in *Saccharomyces cerevisiae*. Biotechnol Bioeng. 85**,** 130-7.

Jiang, H., Wood, K. V., Morgan, J. A., 2005. Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. Appl Environ Microbiol. 71**,** 2962-9.

Jones, J. A., Vernacchio, V. R., Sinkoe, A. L., Collins, S. M., Ibrahim, M. H., Lachance, D. M., Hahn, J., Koffas, M. A., 2016. Experimental and computational optimization of an *Escherichia coli* co-culture for the efficient production of flavonoids. Metab Eng. 35**,** 55-63.

Jullesson, D., David, F., Pfleger, B., Nielsen, J., 2015. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. Biotechnol Adv. 33**,** 1395-402.

Kamei, Y., Hayakawa, T., Urata, J., Uchiyama, M., Shibui, K., Kim, K., Kudo, Y., Okawa, M., 2000. Melatonin treatment for circadian rhythm sleep disorders. Psychiatry Clin Neurosci. 54**,** 381-2.

Kim, B., Cho, B. R., Hahn, J. S., 2014. Metabolic engineering of *Saccharomyces cerevisiae* for the production of 2-phenylethanol via Ehrlich pathway. Biotechnol Bioeng. 111**,** 115-24.

Koopman, F., Beekwilder, J., Crimi, B., Van Houwelingen, A., Hall Robert, D., Bosch, D., Van Maris Antonius, J., Pronk Jack, T., Daran, J.-M., 2012. *De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. Microb Cell Fact. 11**,** 155.

Lee, K., Hahn, J. S., 2013. Interplay of Aro80 and GATA activators in regulation of genes for catabolism of aromatic amino acids in *Saccharomyces cerevisiae*. Mol Microbiol. 88**,** 1120-34.

Li, M., Kildegaard, K. R., Chen, Y., Rodriguez, A., Borodina, I., Nielsen, J., 2015. *De novo* production of resveratrol from glucose or ethanol by engineered *Saccharomyces cerevisiae*. Metab Eng. 32**,** 1-11.

Lu, R., Lu, F., Chen, J., Yu, W., Huang, Q., Zhang, J., Xu, J., 2016. Production of diethyl terephthalate from biomass-derived muconic acid. Angew Chem Int Ed Engl. 55**,** 249-53.

Luttik, M. A. H., Vuralhan, Z., Suir, E., Braus, G. H., Pronk, J. T., Daran, J. M., 2008. Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: Quantification of metabolic impact. Metab Eng. 10**,** 141-153.

Mannhaupt, G., Stucka, R., Pilz, U., Schwarzlose, C., Feldmann, H., 1989. Characterization of the prephenate dehydrogenase-encoding gene, TYR1, from *Saccharomyces cerevisiae*. Gene. 85**,** 303-311.

Minami, H., Kim, J. S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., Sato, F., 2008. Microbial production of plant benzylisoquinoline alkaloids. Proc Natl Acad Sci USA. 105**,** 7393-8.

Moxley, J. F., Jewett, M. C., Antoniewicz, M. R., Villas-Boas, S. G., Alper, H., Wheeler, R. T., Tong, L., Hinnebusch, A. G., Ideker, T., Nielsen, J., Stephanopoulos, G., 2009. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. Proc Natl Acad Sci USA. 106**,** 6477.

Niu, W., Draths, K. M., Frost, J. W., 2002. Benzene-free synthesis of adipic acid. Biotechnol Prog. 18**,** 201-11.

Oud, B., Flores, C. L., Gancedo, C., Zhang, X., Trueheart, J., Daran, J. M., Pronk, J. T., van Maris, A. J., 2012. An internal deletion in MTH1 enables growth on glucose of pyruvate-decarboxylase negative, non-fermentative *Saccharomyces cerevisiae*. Microb Cell Fact. 11**,** 131.

Pandal, N., Global markets for flavors and fragrances. BCC Research, Wellesley, MA, 2014.

Pasquali, G., Porto, D. D., Fett-Neto, A. G., 2006. Metabolic engineering of cell cultures versus whole plant complexity in production of bioactive monoterpene indole alkaloids: recent progress related to old dilemma. J Biosci Bioeng. 101**,** 287-96.

Picataggio, S., Beardslee, T., 2012. Biological methods for preparing adipic acid. Patent number US8241879 B2.

Priefert, H., Rabenhorst, J., Steinbuchel, A., 2001. Biotechnological production of vanillin. Appl Microbiol Biotechnol. 56**,** 296-314.

Ro, D. K., Ehlting, J., Douglas, C. J., 2002. Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. Plant Physiol. 130**,** 1837-51.

Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., Nielsen, J., 2015. Establishment of a yeast platform strain for production of *p*-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. Metab Eng. 31**,** 181-188.

Rodriguez-Naranjo, M. I., Torija, M. J., Mas, A., Cantos-Villar, E., Garcia-Parrilla Mdel, C., 2012. Production of melatonin by *Saccharomyces* strains under growth and fermentation conditions. J Pineal Res. 53**,** 219-24.

Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P., Werck-Reichhart, D., 2001. CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J Biol Chem. 276**,** 36566.

Shin, S. Y., Jung, S. M., Kim, M. D., Han, N. S., Seo, J. H., 2012. Production of resveratrol from tyrosine in metabolically engineered *Saccharomyces cerevisiae*. Enzyme Microb Technol. 51**,** 211-6.

Sonawane, B., Bayliss, D., Valcovic, L., Chen, C., Rodan, B., Farland, W., 2000. Carcinogenic effects of benzene–a status update and research needs to improve risk assessments: US EPA perspective. Environmental Protection Agency. J Toxicol Environ Health A. 61**,** 471-2.

Stark, D., Kornmann, H., Münch, T., Sonnleitner, B., Marison, I. W., Von Stockar, U., 2003. Novel type of in situ extraction: Use of solvent containing microcapsules for the bioconversion of 2-phenylethanol from L-phenylalanine by *Saccharomyces cerevisiae*. Biotechnol Bioeng. 83**,** 376-385.

Stark, D., Münch, T., Sonnleitner, B., Marison, I. W., Stockar, U. V., 2002. Extractive bioconversion of 2-phenylethanol from L-phenylalanine by *Saccharomyces cerevisiae*. Biotechnol Prog. 18**,** 514-523.

Suastegui, M., Guo, W., Feng, X., Shao, Z., 2016a. Investigating strain dependency in the production of aromatic compounds in *Saccharomyces cerevisiae*.

Suastegui, M., Matthiesen, J. E., Carraher, J. M., Hernandez, N., Rodriguez Quiroz, N., Okerlund, A., Cochran, E. W., Shao, Z., Tessonnier, J.-P., 2016b. Combining metabolic engineering and electrocatalysis: application to the production of polyamides from sugar. Angew Chem Int Ed Engl. 55**,** 2368-73.

Sun, X., Lin, Y., Yuan, Q., Yan, Y., 2015. Precursor-directed biosynthesis of 5-hydroxytryptophan using metabolically engineered *E. coli*. ACS Synth Biol. 4**,** 554-8.

Sydor, T., Schaffer, S., Boles, E., 2010. Considerable increase in resveratrol production by recombinant industrial yeast strains with use of rich medium. Appl Environ Microbiol. 76**,** 3361-3.

Szczebara, F. M., Chandelier, C., Villeret, C., Masurel, A., Bourot, S., Duport, C., Blanchard, S., Groisillier, A., Testet, E., Costaglioli, P., Cauet, G., Degryse, E., Balbuena, D., Winter, J., Achstetter, T., Spagnoli, R., Pompon, D., Dumas, B., 2003. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. Nat Biotechnol. 21**,** 143-9.

Trantas, E., Panopoulos, N., Ververidis, F., 2009. Metabolic engineering of the complete pathway leading to heterologous biosynthesis of various flavonoids and stilbenoids in *Saccharomyces cerevisiae*. Metab Eng. 11**,** 355-66.

Trenchard, I. J., Siddiqui, M. S., Thodey, K., Smolke, C. D., 2015. *De novo* production of the key branch point benzylisoquinoline alkaloid reticuline in yeast. Metab Eng. 31**,** 74-83.

Vannelli, T., Wei Qi, W., Sweigard, J., Gatenby, A. A., Sariaslani, F. S., 2007. Production of *p*-hydroxycinnamic acid from glucose in *Saccharomyces cerevisiae* and *Escherichia coli* by expression of heterologous genes from plants and fungi. Metab Eng. 9**,** 142-51.

Wang, Y., Halls, C., Zhang, J., Matsuno, M., Zhang, Y., Yu, O., 2011. Stepwise increase of resveratrol biosynthesis in yeast *Saccharomyces cerevisiae* by metabolic engineering. Metab Eng. 13**,** 455-63.

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., Li, Y., Dowle, A. A., Cartwright, J., Bates, R., Ashford, D., Thomas, J., Walker, C., Bowser, T. A., Graham, I. A., 2015. Plant science. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. Science. 349**,** 309-12.

Zhang, H., Pereira, B., Li, Z., Stephanopoulos, G., 2015. Engineering *Escherichia coli* coculture systems for the production of biochemical products. Proc Natl Acad Sci USA. 112**,** 8266-8271.

Zhang, J., Wu, C., Sheng, J., Feng, X., 2016. Molecular basis of 5-hydroxytryptophan synthesis in *Saccharomyces cerevisiae*. Mol Biosyst.

Zhu, C., Byers, K. J., McCord, R. P., Shi, Z., Berger, M. F., Newburger, D. E., Saulrieta, K., Smith, Z., Shah, M. V., Radhakrishnan, M., Philippakis, A. A., Hu, Y., De Masi, F., Pacek, M., Rolfs, A., Murthy, T., Labaer, J., Bulyk, M. L., 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. 19**,** 556-66.

Ziegler, J., Facchini, P. J., Geissler, R., Schmidt, J., Ammer, C., Kramell, R., Voigtlander, S., Gesell, A., Pienkny, S., Brandt, W., 2009. Evolution of morphine biosynthesis in opium poppy. Phytochemistry. 70**,** 1696-707.

CHAPTER 3

COMBINING METABOLIC ENGINEERING AND ELECTROCATALYSIS: APPLICATION TO THE PRODUCTION OF POLYAMIDES FROM SUGAR

A modified version of the manuscript published in Angewandte Chemie International Edition

Authors: Miguel Suástegui[†,1], John E. Matthiesen[†,1], Jack M. Carraher[1], Nacu Hernandez[1], Natalia Rodriguez Quiroz[1], Adam Okerlund[1], Eric W. Cochran[1], Zengyi Shao[1,*], and Jean-Philippe Tessonnier[1,*]

[1]Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

[†]Joint first authors
[*]Corresponding authors

Abstract

Biorefineries aim to convert biomass to a spectrum of products ranging from biofuels to specialty chemicals. To achieve economically sustainable conversion, it is crucial to streamline the catalytic and downstream processing steps. Here we report a route that integrates bio- and chemical catalysis to convert glucose into bio-based unsaturated nylon 6,6. An engineered strain of *Saccharomyces cerevisiae*, with the highest reported muconic acid titer of 559.5 mg L^{-1} in yeast, was used as the initial biocatalyst to convert glucose into muconic acid. Without any separation, muconic acid was further electrocatalytically hydrogenated to 3-hexenedioic acid with 94% yield, despite the presence of all the biogenic impurities. Bio-based unsaturated nylon 6,6 (unsaturated polyamide 6,6) was finally obtained by polymerization of 3-hexenedioic acid with hexamethylenediamine, demonstrating the integrated design of bio-based polyamides from glucose.

1. Introduction

Biomass emerges as an alternative feedstock to petroleum to grow a more sustainable chemical industry and alleviate the concerns about fossil resources. The transition from fossil to renewable feedstocks is also expected to revitalize the chemical industry by providing building blocks with new functionalities (Hernandez et al., 2014). Since the U.S. Department of Energy's report on top value-added chemicals from biomass (Bozell and Petersen, 2010), extensive research has been carried out to establish biological, chemical, or hybrid pathways for converting cellulosic sugars (Choi et al., 2015; Corma et al., 2007; Schwartz et al., 2014b). Over the past few years, it has become evident that the diversification of building blocks requires the combination of biological and chemical transformations (Corma et al., 2007; Linger et al., 2014; Nikolau et al., 2008; Schwartz et al., 2014b; Vardon et al., 2015; Vennestrøm et al., 2010), *i.e.* biomass is first biologically converted using microbes to platform molecules that are further diversified using chemical catalysis. However, previous attempts to implement chemical and biological processes have led to low conversion rates due to catalyst deactivation by the residual biogenic impurities (Schwartz et al., 2014a; Zhang et al., 2008).

The ideal biorefinery pipelines, from biomass to the final products, are currently disrupted by a gap between biological conversion and chemical diversification. In this work, we report a strategy to bridge this gap by integrating fermentation and electrocatalytic processing. We illustrate this concept with the conversion of glucose to unsaturated polyamide 6,6 (UPA 6,6). The process entails the fermentation of glucose to muconic acid (MA), followed by electrocatalytic hydrogenation (ECH) to 3-hexenedioic acid (HDA), and subsequent polycondensation with 1,6-hexamethylenediamine (HMDA) to yield the desired UPA 6,6 (**Figure 1**). The overall integration in this study is ensured through the utilization of a

metabolically engineered yeast and substitution of conventional high-pressure hydrogenation by direct ECH without separation using the broth water, salts, and impurities as an electrolyte and hydrogen source. With this integration, we demonstrate that a bio-based polymer can be produced using a combined bio- and electrocatalytic process.

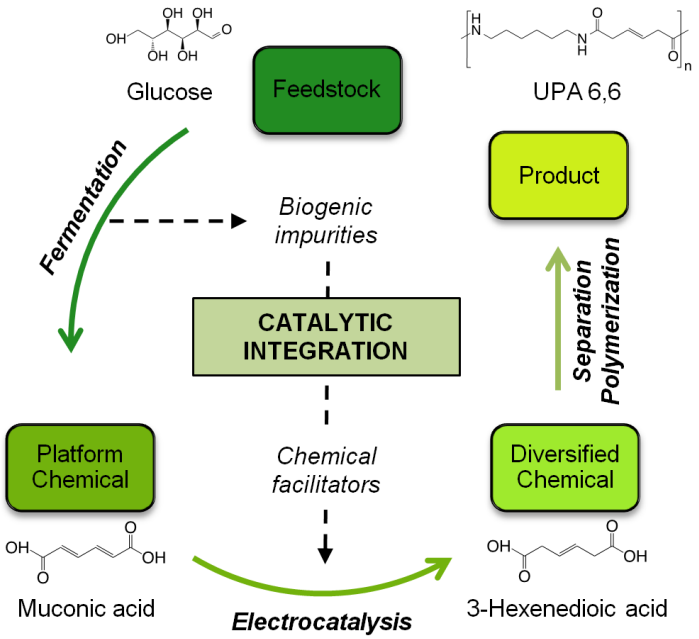


Figure 1: Hybrid conversion of glucose to UPA 6,6. The catalytic integration was enabled by the compatibility of the process parameters. Replacing conventional high-pressure hydrogenation by direct ECH promoted a seamless flow between the processes, allowing the use of the broth water, salts, and impurities as an electrolyte and hydrogen source

2. Materials and Methods

2.1. Strains and media

The yeast strains used in this study are listed in **Table 1**. Cells were grown in YPAD (1% yeast extract, 2% peptone, 0.01% adenine, and 2% dextrose) or synthetic complete media (0.5% ammonium sulfate, 0.16% yeast nitrogen without amino acid and ammonium sulfate, 2% or 4% dextrose) lacking the corresponding auxotrophic nutrient (uracil, tryptophan, leucine, histidine, or in combination). *Escherichia coli* BW25141 was cultured in Luria Bertani (LB) media supplemented with 100 g ml^{-1} ampicillin. Bacteria and yeast were cultured in an orbital shaker incubator at 250 rpm, 37 °C and 30 °C, respectively.

2.2. Plasmid and strain construction

The primers used for gene knockout are listed in **Supplementary Table 1**. Gene knockouts were performed by homologous recombination of a functional *ura3* cassette. Selection in SC+Ura+FOA medium, after transformation of a DNA fragment composed of contiguous upstream and downstream homologs to the original target, resulted in colonies with the desired knockout. All the plasmids assembled in this work derive from the pRS shuttle vector series (New England BioLabs, Ipswich, MA) and were constructed following the DNA Assembler technique(Shao et al., 2009). The genes conforming MA pathway were obtained from the previous work in *S. cerevisiae* (Curran et al., 2013) and codon-optimized by GenScript (Piscataway, NJ), including *Podospora 48ccessio* Pa_5_5120 (*AROZ*), encoding 3-dehydroshikimate dehydratase; *Klebsiella pneumoniae AROY*, encoding protocatechuic acid decarboxylase, and *Candida albicans* (*HQD2)*, encoding 1,2-catechol dioxygenase.

Endogenous genes were amplified from *S. cerevisiae* YSG50 genomic DNA. A detailed list of the plasmids assembled in this work is presented in **Table 2**.

2.3. Flux balance analysis.

Flux balance analysis calculations were performed using MATLAB (Mathworks, Natick, MA) and the COBRA toolbox add-in (Schellenberger et al., 2011). The yeast genome-scale model Imm904 was utilized for the analysis (Mo et al., 2009). The heterologous pathway to produce MA was added to the Imm904 model as described elsewhere (Curran et al., 2013). The glucose and oxygen uptake rates were set to 20 and 2 mmol Gdw^{-1} h^{-1}, respectively (Mo et al., 2009). The solution to the linear problem was calculated with the gurobi linear solver. Initially, the linear system was optimized to maximize the production of biomass. This theoretical value was then set as constraint and flux variability analysis (FVA) was carried out (yielding WT scenario). To simulate the production of MA, the flux to biomass was constrained to 20% of the theoretical maximum, and MA was set as the target. FVA was performed with the new constraints (yielding MA scenario) and the flux through each reaction was compared to the WT scenario.

2.4. ARO1 variants construction

The residues K1370 and D1409 of ARO1 were substituted to alanine by PCR. Two different plasmids were assembled, each containing a different *aro1* mutant, namely, pRS426 $aro1_{K1370A}$ and pRS426 $aro1_{D1409A}$. In parallel, the *E. coli* genes *aroB* and *aroD,* encoding the enzymes DHQ synthase and DHQ dehydratase, were also assembled in a pRS426 vector, together, and in combination with the *aro1* mutants.

Table 1: Engineered InvSc1 strains for production of MA.

Strain	Genotype	Plasmid	Parent strain
InvSc1	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	None	None
InvSc1 Δ*pdc1*	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52* Δ*pdc1*	None	InvSc1
InvSc1 *aro1::aroY*	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52* *aro1::aroY*	None	InvSc1
InvSc1 pRS414 *aroY*	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS414 *aroY*	InvSc1
InvSc1 MA1	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1*	InvSc1
InvSc1 MA2	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 WT *aro1*	InvSc1 MA1
InvSc1 MA3	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{K1370A}$	InvSc1 MA1
InvSc1 MA4	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{D14090A}$	InvSc1 MA1
InvSc1 MA5	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 *hdqs-hdqd*	InvSc1 MA1
InvSc1 MA6	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{K1370A}$-*hdqs-hdqd*	InvSc1 MA1
InvSc1 MA7	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{D14090A}$-*hdqs-hdqd*	InvSc1 MA1
InvSc1 MA8	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52* *aro1::aroY*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{D14090A}$	InvSc1 *aro1::aroY*
InvSc1 MA9	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{D14090A}$-*pc-ppck*	InvSc1 MA1
InvSc1 MA10	*MATa his3Δ1 leu2 trp1-289 ura3-52* *MAT his3Δ1 leu2 trp1-289 ura3-52* Δ*pdc1*	pRS425 Sc MA pRS413 $aro4_{K229L}$-*tkl1* pRS426 $aro1_{D14090A}$ –*pc-ppck*	InvSc1 Δ*pdc1*

Table 2: Plasmids constructed harboring the genetic platforms for MA production.

Production plasmids	Genetic elements	Description
pRS413	Single-copy vector for use in *S. cerevisiae* for selection in media lacking histidine.	
pRS425	Multiple copy vector for use in *S. cerevisiae* for selection in media lacking leucine or uracil, respectively.	
pRS426		
pRS425 Sc MA	pRS425- PYK1p-*AROZ* / GPD1p-*HQD2* / TEF1p-*AROY*	Plasmid containing the heterologous MA pathway
pRS413 *aro4*$_{K229L}$-*tkl1*	pRS413- TP1p-*aro4*$_{K229L}$-THD1$_t$ / ADH1p-*TKL1*-ADH1t	Plasmid harboring the mutant *aro4* gene to remove tyrosine feedback inhibition. It also contains the *tkl1* gene to increase the flux in the non-oxidative portion of pentose phosphate pathway
pRS426 *aro1*	pRS426- GPDp-*ARO1*-PYK1t	Plasmid harboring the pentafunctional gene *ARO1* and its mutant variants *aro1K1370A* and *D1409A*, to stop the conversion of DHS to shikimic acid.
pRS426 *aro1*$_{K1370A}$	pRS426- GPDp-*aro1*$_{K1370A}$-PYK1t	
pRS426 *aro1*$_{D1409A}$	pRS426 GPDp-*aro1*$_{D1409A}$-PYK1t	
pRS426 *dhqs-dhqd*	pRS426 TEF1p-*dhqs*-HXT7t / PYKp-*dhqd*-ADH2*t*	Plasmid containing the *E.coli* genes dehydroquinoate synthase and dehydroquinoate dehydratase alone and in combination with the mutant versions of *ARO1*.
pRS426 *aro1*$_{K1370A}$-*dhqs-dhqd*	pRS426- GPDp-*aro1*$_{K1370A}$-PYK1t / TEF1p-*dhqs*-HXT7t / PYKp-*dhqd*-ADH2*t*	
pRS426 *aro1*$_{D1409A}$-*dhqs-dhqd*	pRS426 GPDp-*aro1*$_{D1409A}$-PYK1t / TEF1p-*dhqs*-HXT7t / PYKp-*dhqd*-ADH2*t*	
pRS426 *aro1*$_{D1409A}$-*PC-PPCK*	pRS426 GPDp-*aro1*$_{D1409A}$-PYK1t / TEF1p-*PC*-HXT7t / PYKp-*PPCK*-ADH2*t*	(Plasmid harboring the mutant *aro1* gene, as well as pyruvate carboxylase and phosphoenolpyruvate carboxykinase for pyruvate recycling to PEP)

2.5. *In vivo* PCA decarboxylase activity assay

The activity of PCA decarboxylase was assessed at different levels of dissolved oxygen. Three different oxygen scenarios were tested: a) aerobic b) oxygen-limited, and c) anaerobic. The oxygen-limited and anaerobic cultures were carried out in 100 Ml serum bottles containing 60 Ml media with shaking at 150 rpm and 30 °C. The oxygen-limited flasks were covered with aluminum foil, whereas the anaerobic cultures were cap-sealed. A plasmid containing only the aroY cassette was assembled in pRS414 plasmid and transformed into strain InvSc1. After 24 hours of growth, the cultures were spiked with 1 Mm PCA. After 18 hours, samples were taken to measure the conversion of PCA to catechol.

2.6. Small-scale and bioreactor fermentations

An Applikon Biotechnology mini-fermenter system (Netherlands) equipped with a 250 Ml vessel was implemented for scale-up fermentation to optimize MA production with strain InvSc1 MA4. 4 Ml of initial culture was used as a seed to inoculate the fermenter at a starting OD_{600} of 0.2 (working volume was 200 Ml). Temperature was maintained at 30 °C, the agitation was set to 500 rpm, and the percentage of dissolved oxygen was controlled by pumping air. All the parameters were manipulated with a built-in process control software *my*-control. Control of Ph was not required for these experiments. The BioXpert Lite software was used for online monitor and data acquisition. All cultures were carried out for 5 days and samples were collected every 24 hours.

2.7. Metabolite detection

An ACQUITY Ultra Performance Liquid Chromatography from Waters (Millford, MA) equipped with a BEH-C18 column (Waters) and a PDA detector was used for detection of MA and the pathway intermediates from the fermentation samples. Samples were prepared by

centrifugation of 1 Ml of fermentation broth at 5,000 rpm for 5 minutes. The supernatant was diluted 10 times in nanopure water and filtered with a 2 μm syringe filter. The mobile phase consisted of 0.2% formic acid in nanopurewater (Solution A) and 0.2% formic acid in acetonitrile (Solution B). The method was as follows: 0.6 ml min-1 of 3% Solution A and 97% Solution B for 1.5 minutes, followed by ramping A to 100% (0% B) in 1 minute. From 3 to 3.1 minutes, Solution A was lowered to 3% (97% B) and maintained until minute 5. The column was maintained at 50 °C and the sample reservoir at 4 °C.

Standard solutions for protocatechuic acid (MP Biomedicals, Santa Ana, CA), catechol (Acros Organics, New Jersey, US) and MA (Sigma, St. Louis, MO) were prepared for quantification. Retention times of PCA, catechol, and MA were 0.7 min, 1.1 min, and 1.5 min, respectively. PCA and MA were detected at 259 nm; catechol was detected at 275 nm.

3. Results

Muconic acid is the unsaturated synthetic precursor of adipic acid and terephthalic acid, which are the corresponding monomers constituting Nylon and polyethylene terephthalate (PET), with a total market value greater than $22 billion (Burk et al., 2011; Niu et al., 2002a; Picataggio and Beardslee, 2012). The traditional benzene-derived synthetic routes for adipic acid and terephthalic acid are environmentally unfriendly (Alini et al., 2007), warranting the need to develop a sustainable and green production platform. For large-scale fermentation, yeast is the preferred microbial host in industry with unique economic advantages such as the greater ease in maintaining phage-free culture conditions and by selling biomass byproducts as animal feeds (In et al., 2005; Jeffries, 2006; Nielsen, 2015; Nielsen et al., 2013).

3.1 Establishing muconic acid production in *S. cerevisiae*

Two previous reports showed the heterologous production of MA in *S. cerevisiae* with titers of 1.56 mg L^{-1} (Weber et al., 2012) and 141 mg L^{-1} (Curran et al., 2013). The low production was caused by a combination of low precursor availability, active competing pathway(s), and the existence of rate-limiting enzyme(s). To address these individual issues, we performed flux balance analysis (FBA) to obtain a list of target genes for genetic manipulations (**Figure 2a**). **Figure 2b** depicts the metabolic pathway with the key manipulations for enhancing MA production. The three genes previously characterized in yeast to produce MA from 3-dehydroshikimic acid (DHS), namely *AroZ* from *Podospora 54ccessio*, *AroY* from *Klebsiella pneumoniae*, and HQD2 from *Candida albicans,* were cloned in a multiple-copy plasmid. Furthermore, the tyrosine-insensitive DAHP synthase ($ARO4_{K229L}$) was overexpressed to remove feedback inhibition caused by the aromatic amino acids in the growth medium. Following our FBA analysis and previous reports to increase the carbon flux into the aromatic amino acid biosynthetic pathway (Cui et al., 2014; Li et al., 1999; Luttik et al., 2008), the transketolase gene (*TKL1*) was overexpressed to increase the pool of the precursor erythrose-4-phosphate (E4P).

Initial fermentations produced 132 mg L^{-1} MA (strain InvSc1 MA1* in **Figure 3**), similar to the titer of the previously highest yeast producer MuA12, but the yield was increased almost 2-fold. We reasoned that this increase is mostly due to the intrinsic difference in the genetic backgrounds of the host strains. Strain dependency in metabolic engineering has also been reported previously (Du et al., 2012b; Kummel et al., 2010). In this case the strain InvSc1 is a diploid whereas the strain MuA12 derives from the haploid *S. cerevisiae* BY4741.

3.2 Improving the synthetic flux towards DHS via a novel mutant

To further force the carbon flow towards DHS, the flux through the two initial reactions in the aromatic amino acid pathway needs to be increased (**Figure 2b**). Suggested by the FBA (**Figure 2a**), the flux through the DHQ dehydratase had to be increased 60-fold to maximize MA production. In yeast, these two reactions are catalyzed by the pentafunctional ARO1 enzyme (Braus, 1991), unlike the self-standing SA pathway enzymes from bacteria and plants. By sequence alignment of ARO1 with the characterized SA dehydrogenases from various species (**Figure 4**), it was concluded that the residues K1370 and D1409 could potentially serve as catalytic residues.

A panel of plasmids was created to enhance the flux towards DHS and avoid conversion into SA. Overexpression of the mutant $ARO1_{D1409A}$ (InvSc1 MA4) increased the production of MA to 235 mg L^{-1} (strain InvSc1 MA4), confirming that the residue D1409 is essential in the catalytic activity of the SA dehydrogenase subunit of ARO1. To ensure that no carbon was being diverted to SA, we deleted both copies of *aro1* (InvSc1 MA8 in **Figure 3**), but the MA production unexpectedly decreased to 25 mg L^{-1} and the strain's fitness was also affected even when aromatic amino acids were supplemented.

To increase the availability of phosphoenolpyruvate (PEP), we overexpressed *PC* (pyruvate carboxylase) and *PPCK* (phosphoenolpyruvate carboxykinase) in strain InvSc1-MA4, as well as in the strain InvSc1Δ*pdc1* (resulting in strains InvSc1 MA9 and InvSc1 MA10, respectively). However, the MA titers decreased by around 60% and 74%, respectively (**Figure 3**). Recirculation of pyruvate to PEP has been successfully applied in *Escherichia coli* to increase the yield of aromatic compounds. (Patnaik and Liao, 1994a; Zhang et al., 2015a) In *S. cerevisiae*, the failure of this strategy might be attributed to the inaccuracy of FBA

modeling to predict the compartmentalization fluxes, enzyme kinetics, and the metabolic burden on the cells caused by overexpression of highly regulated genes.

3.3 Controlling oxygen to alleviate the PCA bottleneck

Despite the aforementioned genetic manipulations, the production of MA was still limited by the accumulation of the intermediate protocatechuic acid (PCA). The enzyme PCA decarboxylase is known to be oxygen sensitive (He and Wiegel, 1996; Weber et al., 2012); we observed that the conversion of PCA to catechol increased more than 2-fold when growing the cells in an oxygen-limited culture (**Figure 5a**). This improved the ratio of MA to PCA to 2.5, which is almost 5-fold higher than the previously highest MA yeast producer. The highest MA titer reached 559.5 mg L^{-1}, representing a 4-fold improvement in both titer and yield over strain MuA12 (**Figure 5b**).

The MA platform consisted in establishing the heterologous pathway in a diploid yeast strain, with overexpression of the novel mutant $ARO1_{D1409A}$, and alleviation of the PCA bottleneck with a controlled oxygen environment in the fermentation.

The yield of 14 mg_{MA} $g_{glucose}^{-1}$ represents the highest one among the aromatic amino acid-based metabolites that have been produced in yeast in batch fermentation. Given the strong industrial interest in muconic acid (Horwitz et al., 2015; Rogers et al., 2013), substantial improvement should be expected in the next several years. Mutagenesis of AroY to remove its oxygen sensitivity (Stapleton and Swartz, 2010) and global genetic perturbations (DiCarlo et al., 2013; Si et al., 2015) coupled with molecular sensors (Chugani et al., 1998; Ezezika et al., 2006) are potential strategies that will enable yeast to reach higher yields as observed in the *E. coli* platforms (Niu et al., 2002a; Zhang et al., 2015a)

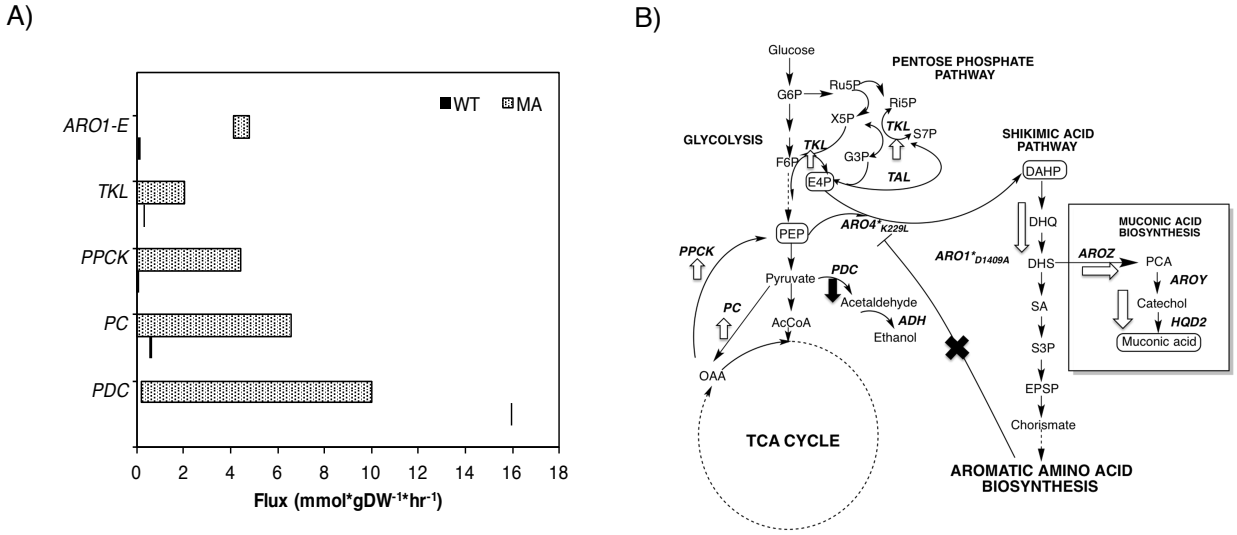


Figure 2: Metabolic engineering rationale for overproduction of MA in *S. cerevisiae*. A) *In silico* modeling of MA in *S. cerevisiae*. Flux variability analysis was performed under two different scenarios. WT represents the flux through selected reactions targeting maximal biomass production. Similarly, the MA scenario represents the flux through the selected reactions having maximal MA as target while maintaining 20% of the maximum theoretical biomass. B) Three main strategies were studied: removing the feedback inhibition by aromatic amino acids, increasing the pool of the precursors PEP and E4P, and increasing the pull of carbon into the shikimic acid pathway. Abbreviation of metabolites – G6P, glucose-6-p; F6P, fructose-6-p; Ru5P, ribulose-5-p; Ri5P, ribose-5-p; S7P, sedoheptulose-5-p; X5P, xylulose-5-p; G3P, glyceraldehyde-3-p; AcCoA, acetyl-CoA; PEP, phosphoenolpyruvate; OAA, oxaloacetate; DAHP, 3-Deoxy-D-arabinoheptulosonate 7-p; DHQ, 3-dehydroquinate; DHS, 3-dehydroshikimic acid; SA, shikimate; S3P, shikimate-3-p; EPSP, 5-enolpyruvylshikimate-3-p; PCA, protocatechuic acid. White and black arrows represent up and downregulated reactions, respectively.

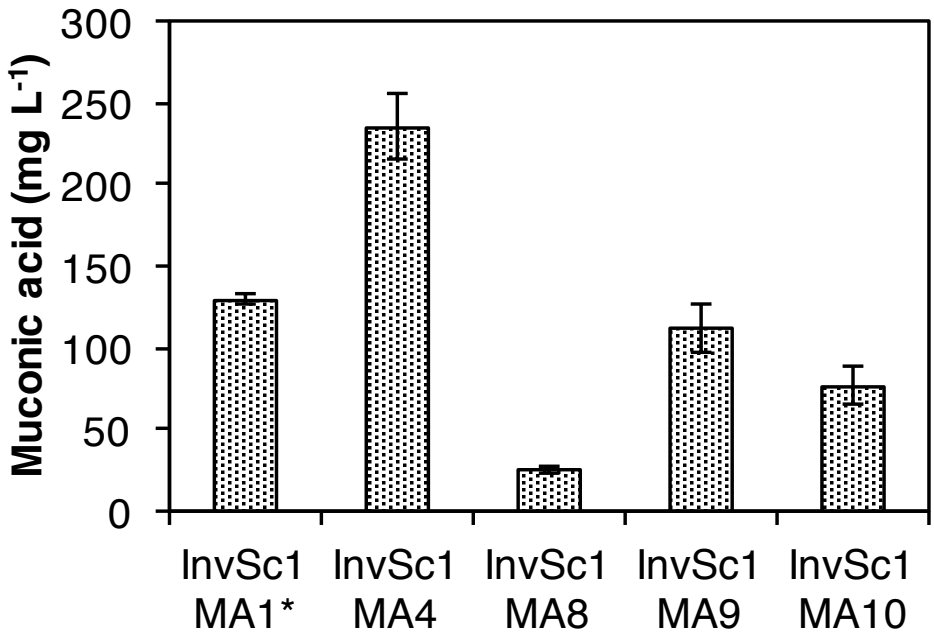


Figure 3: Characterization of MA production in engineered InvSc1 strains. MA accumulation in strains grown in glass tubes containing 15 Ml of selective media. Maximum MA acid was observed at 96 h of aerobic fermentation.

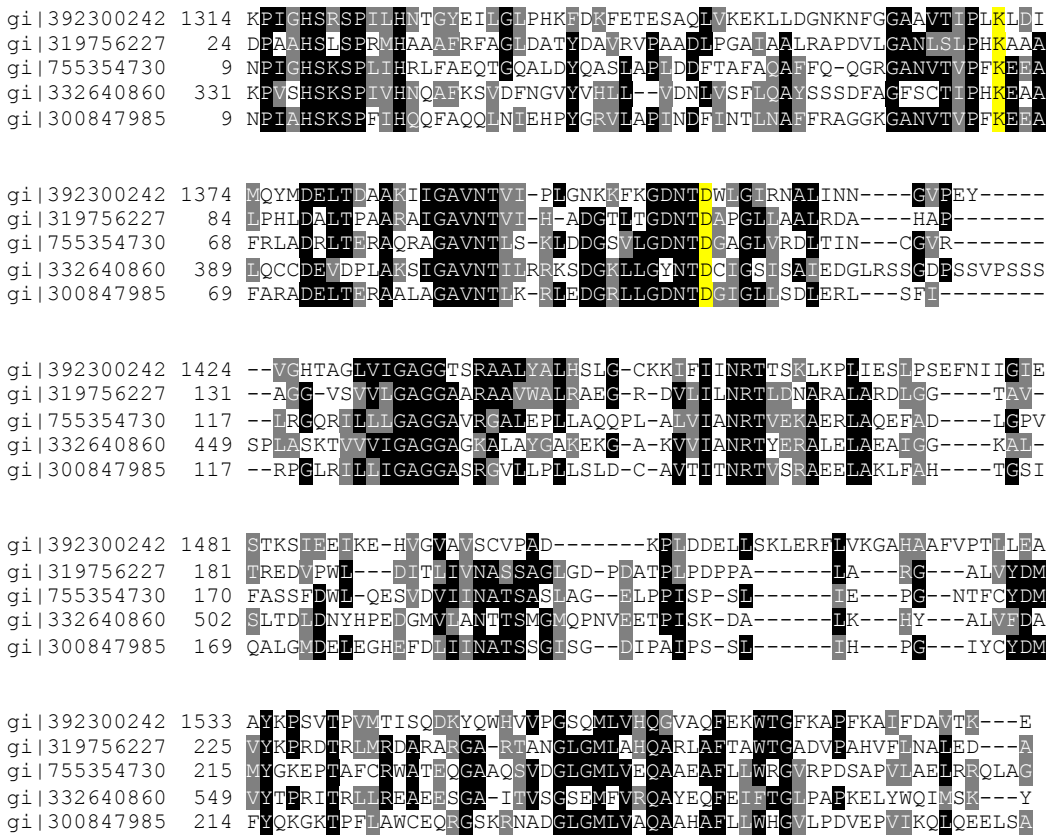


Figure 4: Sequence alignment of ARO1D with DHS dehydrogenases from various species. The ARO1 pentafunctional enzyme from *S. cerevisiae* (gi: 392300242) was aligned with sequences from *Deinococcus maricopensis* (gi: 319756227), *Pseudomonas viridiflava* (gi: 755354730), *Arabidopsis thaliana* (gi: 332640860), and *E. coli* (gi: 300847985). The residues K1370 and D1409 in ARO1 (yellow) showed 100% homology among all the species.

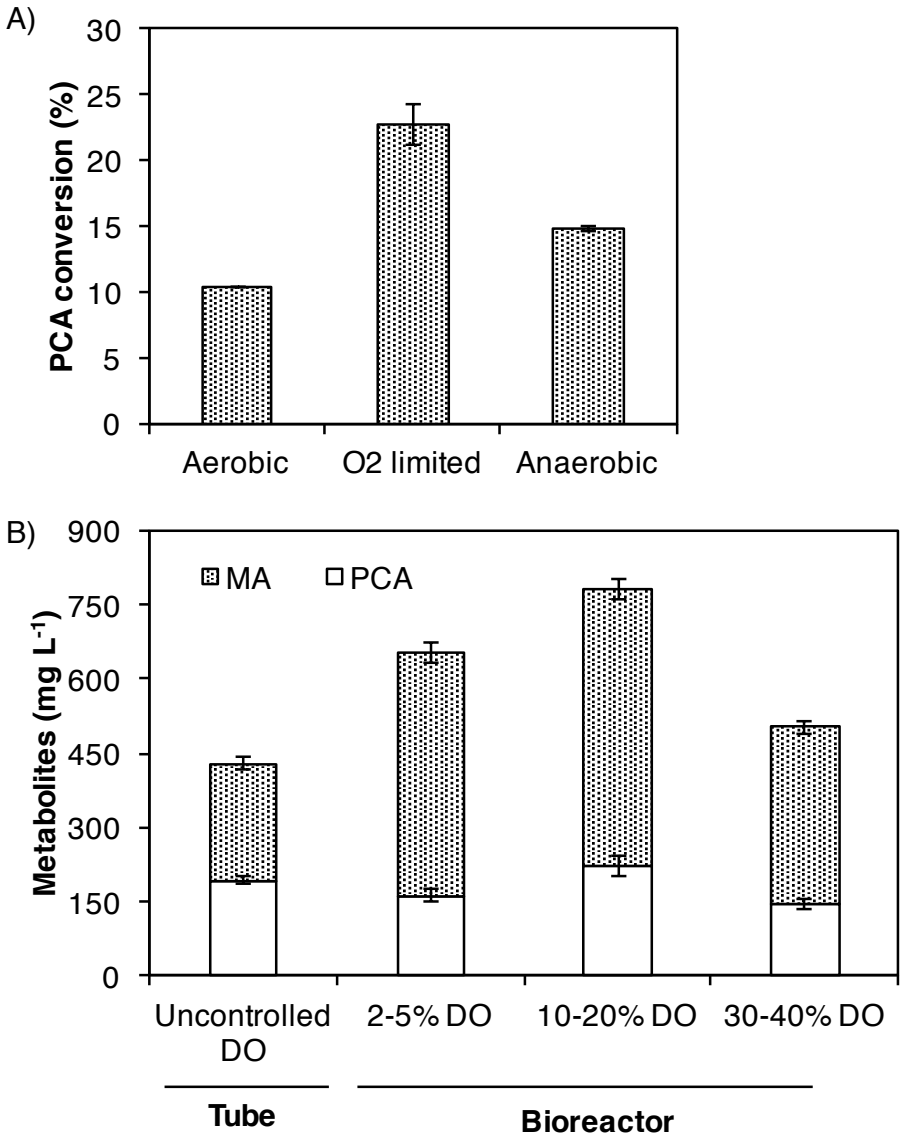


Figure 5: Variation of fermentation conditions. PCA decarboxylase activity assay under different oxygen environments. A) The strain InvSc1 pRS414 aroY was cultured in three different oxygen conditions and spiked at 24 h with 1 Mm PCA. The conversion percentages were calculated based on the samples collected at 18 h after PCA supplementation. B) Strain InvSc1MA4 in mini reactor fermentations with controlled dissolved oxygen. The highest MA accumulation was observed when the dissolved oxygen was maintained between 10 to 20% during the first 24 h. After this period, the dissolved oxygen was set to 20% until the end of the fermentation (4 days).

3.4 Electrochemical hydrogenation

The fermentation broth was subsequently hydrogenated in a three-electrode electrochemical cell (**Figure 6**). Electrocatalysis was preferred over conventional high-pressure hydrogenation as hydrogen is produced *in situ* by water splitting, the reaction occurs under ambient temperature and pressure, and the charge on the electrode surface can mitigate poisoning (Chapuzet et al., 1998). In this configuration, hydrogen production and MA hydrogenation take place simultaneously at the cathode (Eq. 1-2), allowing a seamless ECH.

$$H^{(+)} + e^- + * \rightarrow H^* \text{ [Eq. (1)]}$$

$$2H^* + C_6H_6O_4 + * \rightarrow C_6H_8O_4 \text{ [Eq. (2)]}$$

Lead (Pb) was chosen as a catalyst based on its Earth abundance, low cost, common use in industrial electrosynthetic applications, metallic state under cathodic potentials, and stability in the presence of sulfur (Joyner et al., 1978; Pletcher and Walsh, 1993). The expected resistance to impurities allowed us to significantly simplify the hydrogenation reaction by placing the fermentation broth directly in the electrochemical reactor. The broth contained whole yeast cells, unspent salts, and biogenic impurities coming from cellular metabolism and lysis. The ECH was then allowed to proceed at room temperature and atmospheric pressure for 1 hour by applying a potential of -1.5 V vs. Ag/AgCl on a 10 cm^2 lead rod, resulting in 95% MA conversion with 81% selectivity to HDA. To assess the stability of the catalyst in the fermentation media (in the presence of all potential poisons), five successive one-hour electrocatalytic batch reactions were performed (**Figure 6b**). Notably, no signs of deactivation were observed and leaching of the catalyst into the solution was minimal at 6.5 ± 0.4 ppm as determined by elemental analysis.

To further increase the yield of HDA, the effects of Ph and applied voltage were investigated independently. A model solution of pure MA dissolved in a potassium sulfate/sulfuric acid electrolyte was used to accurately control ionic strength and to maintain a constant ionic conductivity. Acidic conditions favored the selective formation of HDA, especially for reaction times below 30 min (**Supplementary Figure 1**). Further ^{1}H nuclear magnetic resonance (NMR) analysis of a HDA model solution after ECH revealed that the observed decrease in selectivity as the reaction proceeded was due to the formation of decomposition products through secondary reactions and not through the formation of additional hydrogenation products *e.g.* adipic acid.

Conditions optimized with the model solutions were found to also enhance the hydrogenation of the fermentation broth (**Figure 6c**). Notably, when the Ph of the solution was fixed at 2.0, the selectivity towards HDA became 98 ± 4% at 96 ± 2% MA conversion. It is worth noting that the yield achieved for the unpurified broth was actually higher than for the model solution (94% vs. 77%). While catalyst poisoning is a common issue for most of the hydrogenation reactions catalyzed by precious metals (Schwartz et al., 2014a; Zhang et al., 2008), it appears that the broth's impurities were beneficial in the present work as it prevented the formation of decomposition products during the ECH.

3.5 Polymerization of HDA

In the pursuit of a fully integrated process, from glucose to a commercially viable product, HDA was separated from the fermentation broth by vacuum evaporation, activated carbon filtration, and crystallization. High purity (98%) of HDA at a yield of 67% was obtained and subjected to the final step of polymerization. The corresponding saturated nylon 6,6 was synthesized using adipic acid and HMDA in an attempt to compare the conventional

petrochemical-based nylon 6,6 and the bio-based UPA 6,6 polymers. The obtained UPA 6,6 consisted of a transparent, partly crystalized material with comparable physical and chemical properties to petrochemical nylon (**Table 3, Figure 7, Supplementary Table 2**). Polymers based on blends of HDA and adipic acid were also achieved (**Figure 7a**) to enable different levels of 63ccession63. These HDA-containing nylon materials offer precious grafting sites to custom-tailor the existing nylon grades with desirable properties, for example crosslinkability, paintability, flame retardance, etc.

Table 3: The properties of Nylon 6,6 and UPA 6,6. M_n indicates the number average molecular weight, PDI: polydispersity, Q*: primary diffraction peak, G′: storage modulus, G″: loss modulus, and G^*_c: crossover modulus.

Property	Nylon 6,6	UPA 6,6
M_n (Da)	17,800	12,200
PDI	2.0	3.36
Melting Temp (°C)	250	60
Q* ($Å^{-1}$)	4.4	4.7
G’ (Mpa)	-	18.9
G’’ (Mpa)	-	6.24
G''_C (°C)	-	60

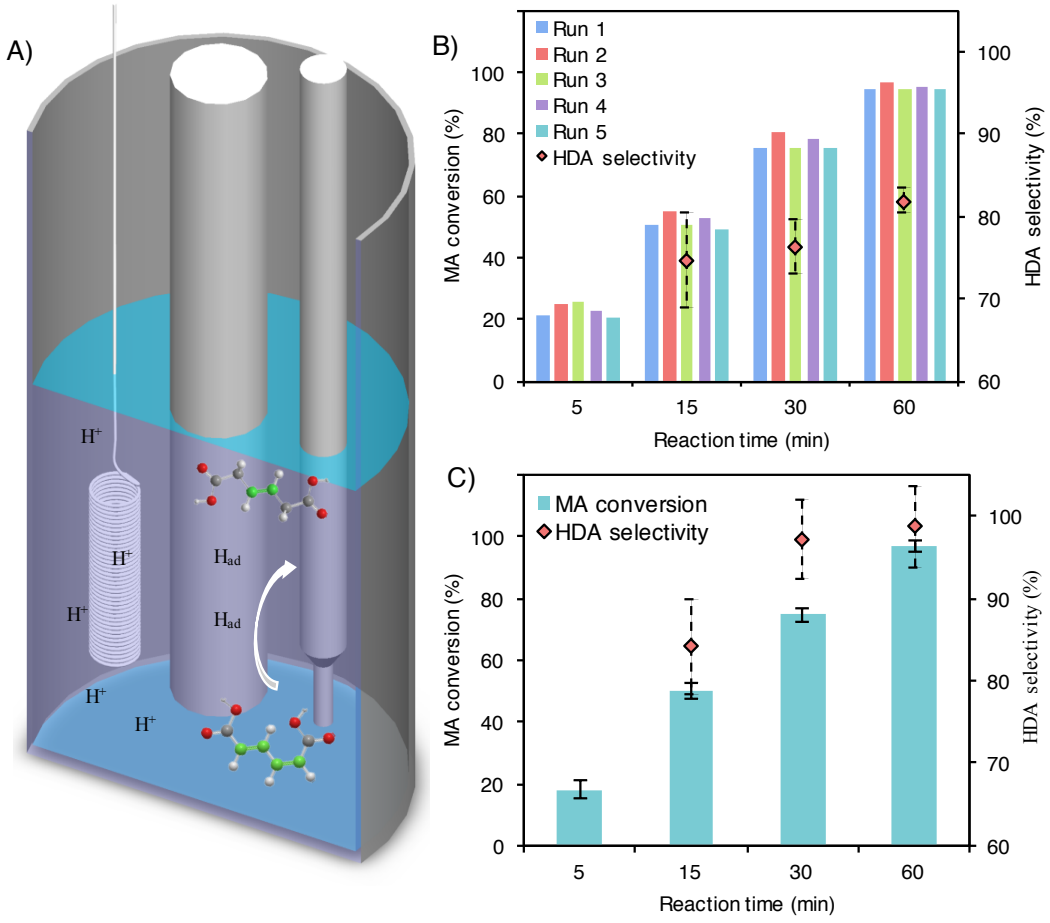


Figure 6: Electrocatalytic hydrogenation of MA to HDA directly in the fermentation broth. The hydrogen necessary for the reaction is generated *in situ* (H_{ad}) at the surface of the Pb electrode. A) Electrocatalytic single-cell reactor for conversion of MA to HDA. Atoms in the molecules are color-coded: grey: carbon; white: hydrogen; red: oxygen; blue: nitrogen; green: carbon-carbon double bond. The reaction was performed under ambient temperature and pressure using a three-electrode electrochemical cell at -1.5 V vs. Ag/AgCl. B) Conversion of MA and average selectivity to the desired product showed no signs of catalyst deactivation when repeating the reaction five successive times (runs 1-5). C) MA and HDA conversion and selectivity for the ECH of the doped (Ph 2.0, -1.5 V) fermentation broth.

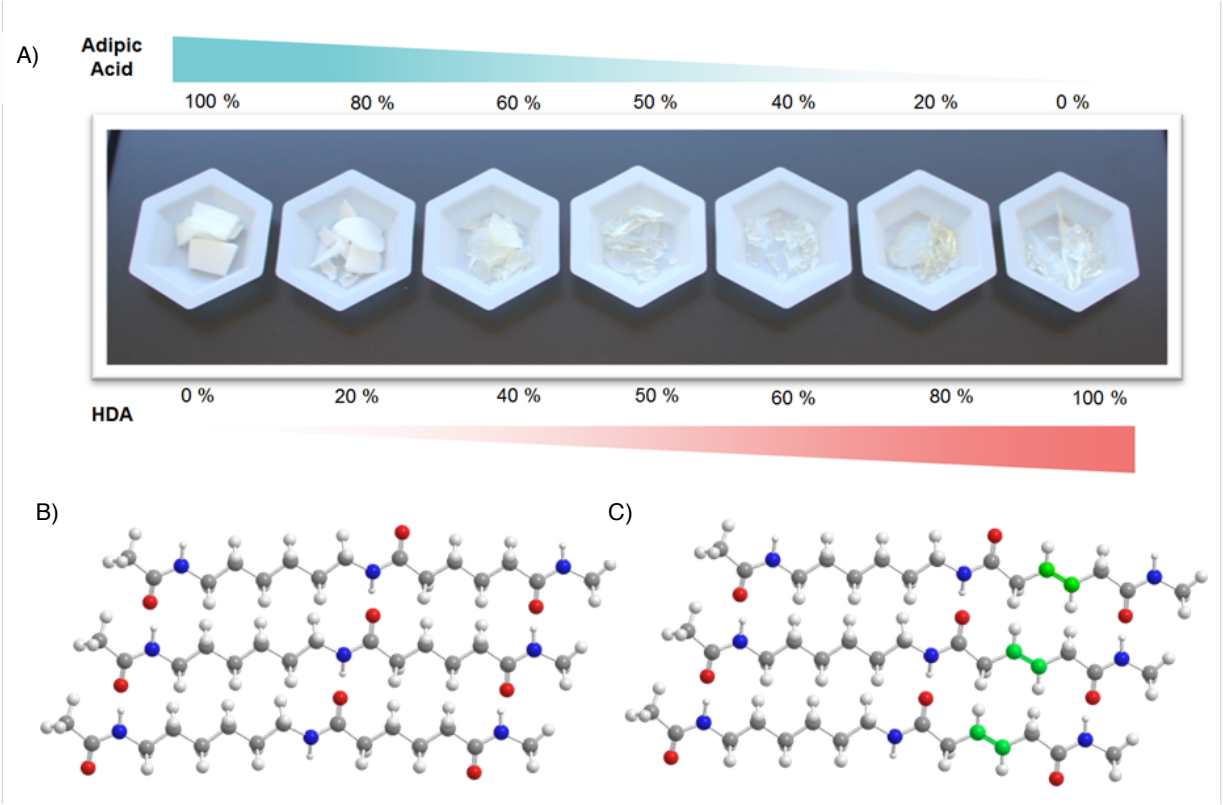


Figure 7: Polymer visualization. A) Polymer blends of adipic acid and HDA. Percentages are based on molar ratios of adipic acid and HDA reacted with a 1:1 mol ratio of HMDA. B) Nylon 6,6 and C) UPA 6,6, for which petroleum-based adipic acid was substituted with HDA; atoms in both molecular structures are color-coded for easier identification. Grey: carbon; white: hydrogen; red: oxygen; blue: nitrogen; green: carbon-carbon double bond

4. Discussion

Development of diverse microbial cell factories as biorefineries has been thriving in the past ten years, with a steady growing number of bio-based chemicals reaching production at commercial scale (Choi et al., 2015; Straathof, 2014). The biosynthetic routes of many key monomers for the polymer industry, including diols, dicarboxylic acids, ω-amino acids, aromatics, diamines, and hydroxyl acids, have been established at various production levels via implementing metabolic engineering and synthetic biology strategies. These monomers were then chemically converted to polymers such as nylon, polyethylene terephthalate (PET), polyurethane, and polystyrene (Chung et al., 2015). Many companies (*i.e.* DSM, BASF, Cargill, DuPont, Verdezyne, Myriant, BioAmber, and Genomatica) are actively involved in the development of bio-based monomer production. Some successful cases include the production of succinic acid as a polyamide monomer, and 1,4-butanediol or 1,3-propanediol as polyester monomers (Choi et al., 2015).

Cis, cis-Muconic acid is the unsaturated synthetic precursor of adipic acid and terephthalic acid, which are the corresponding monomers constituting Nylon and PET, with a total market value greater than $22 billion (Burk et al., 2011; Niu et al., 2002; Picataggio and Beardslee, 2012). The traditional benzene-derived synthetic routes for adipic acid and terephthalic acid are environmentally unfriendly, warranting the need to develop a sustainable and green production platform. The design of biological production methods using fermentation as an alternative to petrochemistry addresses this need. For large-scale fermentation, yeast is the preferred microbial host in industry with unique economic advantages such as the greater ease in maintaining phage-free culture conditions and by selling biomass byproducts as animal feeds (In et al., 2005; Nielsen, 2015; Nielsen et al., 2013). Particularly for dicarboxylic acids,

yeast-based fermentation is preferred over *E. coli*-based platform due to yeast's high tolerance to the gradually decreasing Ph occurring during the production of the desired organic acids. In addition, being able to produce and purify these products in their protonic forms (instead as carboxylates) is economically important considering the significant cost saving by avoiding the usage of alkali and acid titrants and the downstream disposal of salt byproducts (*e.g.* gypsum) (Abbott et al., 2009; Xiao et al., 2014). Here we achieved the highest production of MA in *S. cerevisiae*, representing a 4-fold improvement over the previous effort (Curran et al., 2013). The success was attributed to two key innovations, including (1) the introduction of the mutant Aro1 to increase the synthetic flux towards DHS at the same time halting it from being further converted to shikimate, and (2) improving the activity of the oxygen-sensitive AroY by controlling the oxygen supplied during fermentation.

Although the current titer of MA is still relatively low when compared to the *E. coli* platforms (Niu et al., 2002; Zhang et al., 2015), industry *e.g.,* Amyris Biotechnologies (Horwitz et al., 2015) and Myriant (Rogers et al., 2013) has shown strong interest in the yeast production platform. Substantial, improvement should be expected in the next several years. Most recently, a yeast strain was engineered to produce *p*-coumaric acid using fed-batch fermentation at a titer of about 2 g/L (Rodriguez et al., 2015b). Considering that the biosynthesis of *p*-coumaric acid shares the same precursor with the MA pathway, this success evidences the room to improve MA production. Furthermore, the rate-limiting constraint posed by AroY can be approached by applying directed evolution strategies to relieve its oxygen sensitivity (Stapleton and Swartz, 2010). A high throughput strain optimization method established on MA sensors (Chugani et al., 1998; Ezezika et al., 2006) could be used to screen AroY variant library or identify beneficial genotypes on genome scale that could not be easily

predicted by rational design. Screening global genetic perturbation has proven to be an effective strategy for engineering complex phenotypes in yeast; to date, at least 4000 regulatory interactions have been identified to tightly control the primary metabolism in yeast (Herrgard et al., 2004).

Following fermentation, metal-catalyzed hydrogenation represents one of the main strategies to diversify building block molecules. The resistance to impurities allowed us to significantly simplify the hydrogenation reaction by placing the fermentation broth directly in the electrochemical reactor. Under optimized conditions, the selectivity towards HDA reached 98 ± 4% at 96 ± 2% MA conversion. Moreover, the successful synthesis of UPA 6,6 with various ratios of HDA and adipic acid demonstrates the compatibility of the two monomers. The unsaturated polyamide presents the advantage of having an extra double bond in its backbone that can be used to incorporate additional functionality in Nylon 6,6, *e.g.* for tuned hydrophobicity and hydrophilicity. In addition to polyamides, HDA has been previously used as a precursor to generate dodecanedioic acid (Frost et al., 2010), a monomer of Nylon 6,12, and to produce polyester ethers with biomedical applications (Olson et al., 2006; Vardon et al., 2015).

This work presents one strategy to bridge the gap between biological and chemical catalysis in biorefineries. We demonstrated its potential by synthesizing a new family of unsaturated polyamides from sugar and anticipate that this strategy will facilitate the incorporation of fermentation and catalytic hydrogenation for a broad range of reactions. Future efforts will be taken towards (1) scale-up of the hybrid pipeline with a detailed technoeconomic analysis to assess the cost efficiency and (2) ultimate-level integration combining biocatalytic and chemocatalytic reactions into a one-pot process, which represents an attractive solution to

develop economically and ecologically advantageous chemical synthesis schemes in the context of the water, energy, and food nexus (Gröger and Hummel, 2014).

5. Bibliography

Abbott, D. A., Zelle, R. M., Pronk, J. T., van Maris, A. J., 2009. Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. FEMS Yeast Res. 9**,** 1123-36.

Alini, S., Basile, F., Blasioli, S., Rinaldi, C., Vaccari, A., 2007. Development of new catalysts for N2O-decomposition from adipic acid plant. Appl Catal, B. 70**,** 323-329.

Bozell, J. J., Petersen, G. R., 2010. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's " Top 10" revisited. Green Chem. 12**,** 539.

Braus, G. H., 1991. Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway. Microbiol Rev. 55**,** 349-70.

Burk, M. J., Osterhout, R. E., Sun, J., 2011. Semi-synthetic terephthalic acid via microorganisms that produce muconic acid. Patent number US 20110124911 A1.

Chapuzet, J. M., Lasia, A., Lessard, J., 1998. Electrocatalytic Hydrogenation of Organic Compounds. Wiley.

Choi, S., Song, C. W., Shin, J. H., Lee, S. Y., 2015. Biorefineries for the production of top building block chemicals and their derivatives. Metab Eng. 28**,** 223-239.

Chugani, S. A., Parsek, M. R., Chakrabarty, A. M., 1998. Transcriptional repression mediated by LysR-type regulator CatR bound at multiple binding sites. J Bacteriol. 180**,** 2367-2372.

Chung, H., Yang, J. E., Ha, J. Y., Chae, T. U., Shin, J. H., Gustavsson, M., Lee, S. Y., 2015. Bio-based production of monomers and polymers by metabolically engineered microorganisms. Curr Opin Biotechnol. 36**,** 73-84.

Corma, A., Iborra, S., Velty, A., 2007. Chemical routes for the transformation of biomass into chemicals. Chem. Rev. 107**,** 2411-2502.

Cui, Y.-Y., Ling, C., Zhang, Y.-Y., Huang, J., Liu, J.-Z., 2014. Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. Microb Cell Fact. 13**,** 21/1-21/11, 11 pp.

Curran, K. A., Leavitt, J. M., Karim, A. S., Alper, H. S., 2013. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng. 15**,** 55-66.

DiCarlo, J. E., Conley, A. J., Penttila, M., Jantti, J., Wang, H. H., Church, G. M., 2013. Yeast oligo-mediated genome engineering (YOGE). ACS Synth Biol. 2**,** 741-9.

Du, J., Yuan, Y., Si, T., Lian, J., Zhao, H., 2012. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res. 40**,** e142.

Ezezika, O. C., Collier-Hyams, L. S., Dale, H. A., Burk, A. C., Neidle, E. L., 2006. CatM regulation of the benABCDE operon: Functional divergence of two LysR-type paralogs in *Acinetobacter baylyi* ADP1. Appl Environ Microbiol. 72**,** 1749-1758.

Frost, J. W., Millis, J., Tang, Z., Methods for producing dodecanedioic acid and derivatives thereof. Draths Corporation, USA . 2010, pp. 49 pp.

Gröger, H., Hummel, W., 2014. Combining the ‘ two worlds’ of chemocatalysis and biocatalysis towards multi- step one- pot processes in aqueous media. Curr Opin Chem Biol. 19**,** 171-179.

He, Z., Wiegel, J., 1996. Purification and characterization of an oxygen-sensitive, reversible 3,4-dihydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. J Bacteriol. 178**,** 3539-3543.

Hernandez, N., Williams, R. C., Cochran, E. W., 2014. The battle for the "green" polymer. Different approaches for biopolymer synthesis: bioadvantaged vs. bioreplacement. Org Biomol Chem. 12**,** 2834-49.

Herrgard, M. J., Covert, M. W., Palsson, B. O., 2004. Reconstruction of microbial transcriptional regulatory networks. Curr Opin Biotechnol. 15**,** 70-7.

Horwitz, A. A., Walter, J. M., Schubert, M. G., Kung, S. H., Hawkins, K., Platt, D. M., Hernday, A. D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S. S., Newman, J. D., 2015. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Systems. 1**,** 88-96.

In, M.-J., Kim, D. C., Chae, H. J., 2005. Downstream process for the production of yeast extract using brewer's yeast cells. Biotechnol Bioprocess Eng. 10**,** 85-90.

Jeffries, T. W., 2006. Engineering yeasts for xylose metabolism. Curr Opin Biotechnol. 17**,** 320-6.

Joyner, R. W., Kishi, K., Roberts, M. W., 1978. Low energy electron diffraction and electron spectroscopic studies of the oxidation and sulfidation of lead(100) and lead(110) surfaces. Proc R Soc London, Ser A. 358**,** 223-41.

Kummel, A., Ewald, J. C., Fendt, S. M., Jol, S. J., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., Heinemann, M., 2010. Differential glucose repression in common yeast strains in response to HXK2 deletion. FEMS Yeast Res. 10**,** 322-32.

Li, K., Mikola, M. R., Draths, K. M., Worden, R. M., Frost, J. W., 1999. Fed-batch fermentor synthesis of 3-dehydroshikimic acid using recombinant Escherichia coli. Biotechnol Bioeng. 64**,** 61-73.

Linger, J. G., Vardon, D. R., Guarnieri, M. T., Karp, E. M., Hunsinger, G. B., Franden, M. A., Johnson, C. W., Chupka, G., Strathmann, T. J., Pienkos, P. T., Beckham, G. T., 2014. Lignin valorization through integrated biological funneling and chemical catalysis. Proc Natl Acad Sci U S A Early Ed. 111**,** 12013-12018.

Luttik, M. A. H., Vuralhan, Z., Suir, E., Braus, G. H., Pronk, J. T., Daran, J. M., 2008. Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: Quantification of metabolic impact. Metab Eng. 10**,** 141-153.

Mo, M. L., Palsson, B. O., Herrgard, M. J., 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol. 3**,** 37.

Nielsen, J., 2015. Yeast cell factories on the horizon. Science. 349**,** 1050-1.

Nielsen, J., Larsson, C., van Maris, A., Pronk, J., 2013. Metabolic engineering of yeast for production of fuels and chemicals. Curr Opin Biotechnol. 24**,** 398-404.

Nikolau, B. J., Perera, M. A. D. N., Brachova, L., Shanks, B., 2008. Platform biochemicals for a biorenewable chemical industry. Plant J. 54**,** 536-545.

Niu, W., Draths, K. M., Frost, J. W., 2002. Benzene-free synthesis of adipic acid. Biotechnol Prog. 18**,** 201-11.

Olson, D. A., Gratton, S. E. A., DeSimone, J. M., Sheares, V. V., 2006. Amorphous linear aliphatic polyesters for the facile preparation of tunable rapidly degrading elastomeric devices and delivery vectors. J Am Chem Soc. 128**,** 13625-13633.

Patnaik, R., Liao, J. C., 1994. Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. Appl Environ Microbiol. 60**,** 3903-8.

Picataggio, S., Beardslee, T., 2012. Biological methods for preparing adipic acid. Patent number US8241879 B2.

Pletcher, D., Walsh, F., 1993. Corrosion and its control. Industrial Electrochemistry. Springer Netherlands, pp. 481-542.

Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., Nielsen, J., 2015. Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. Metab Eng. 31**,** 181-188.

Rogers, Y., Gong, W., Dole, S., Sillers, R., Gandhi, M., Pero, J., 2013. Production of muconic acid from genetically engineered microorganisms. Patent number WO2013116244 A1.

Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R., Palsson, B. O., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 6**,** 1290-307.

Schwartz, T. J., Johnson, R. L., Cardenas, J., Okerlund, A., Da Silva, N. A., Schmidt-Rohr, K., Dumesic, J. A., 2014a. Engineering catalyst microenvironments for metal-catalyzed hydrogenation of biologically derived platform chemicals. Angew Chem, Int Ed. 53**,** 12718-22.

Schwartz, T. J., O'Neill, B. J., Shanks, B. H., Dumesic, J. A., 2014b. Bridging the chemical and biological catalysis gap: challenges and outlooks for producing sustainable chemicals. ACS Catal. 4**,** 2060-2069.

Shao, Z., Zhao, H., Zhao, H., 2009. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37**,** e16.

Si, T., Luo, Y. Z., Bao, Z. H., Zhao, H. M., 2015. RNAi-assisted genome evolution in *Saccharomyces cerevisiae* for complex phenotype engineering. ACS Synth Bio. 4**,** 283-291.

Stapleton, J. A., Swartz, J. R., 2010. Development of an *in vitro* compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. PLoS One. 5**,** e15275.

Straathof, A. J. J., 2014. Transformation of biomass into commodity chemicals using enzymes or cells. Chem Rev. 114**,** 1871.

Vardon, D. R., Franden, M. A., Johnson, C. W., Karp, E. M., Guarnieri, M. T., Linger, J. G., Salm, M. J., Strathmann, T. J., Beckham, G. T., 2015. Adipic acid production from lignin. Energy Environ Sci. 8**,** 617-628.

Vennestrøm, P. N. R., Christensen, C. H., Pedersen, S., Grunwaldt, J. d., Woodley, J. M., 2010. Next-generation catalysis for renewables: combining enzymatic with inorganic heterogeneous catalysis for bulk chemical production. ChemCatChem. 2**,** 249-258.

Weber, C., Bruckner, C., Weinreb, S., Lehr, C., Essl, C., Boles, E., 2012. Biosynthesis of cis,cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by *Saccharomyces cerevisiae*. Appl Environ Microbiol. 78**,** 8421-30.

Xiao, H., Shao, Z., Jiang, Y., Dole, S., Zhao, H., 2014. Exploiting *Issatchenkia orientalis* SD108 for succinic acid production. Microb Cell Fact. 13**,** 121.

Zhang, H., Pereira, B., Li, Z., Stephanopoulos, G., 2015. Engineering *Escherichia coli* coculture systems for the production of biochemical products. Proc Natl Acad Sci U S A. 112**,** 8266-8271.

Zhang, Z., Jackson, J. E., Miller, D. J., 2008. Effect of biogenic fermentation impurities on lactic acid hydrogenation to propylene glycol. Bioresource Technol. 99**,** 5873-80.

6. Supplementary Information

6.1. Supplementary Methods

6.1.1. Electrocatalytic hydrogenation

The electrochemical studies were conducted in a three-electrode electrochemical cell using a BioLogic VSP-300 potentiostat. The Ag/AgCl in NaCl reference electrode and platinum counter electrode were purchased from BioLogic Science Instruments. The working electrode was purchased from Rotometals (Lead rod, 99.9%). Electrocatalytic hydrogenation experiments were performed in 11 Ml of reacting medium with a stir bar at 700 rpm. During the chronoamperometry experiments, 0.5 Ml samples of the reaction medium were taken at 5 min, 15 min, 30 min, and 60 min for analysis.

Samples were subsequently analyzed either by ultra-performance liquid chromatography (UPLC) or ^{1}H NMR. For NMR analysis, the samples were dried at room temperature, reconstituted in deuterium oxide, and analyzed with a Bruker 600 MHz NMR spectrometer (AVIII600). UPLC analysis was performed with a Waters H-Class Acquity chromatograph equipped with a HSS C18 Column (1.8 μm, 2.1 x 100 mm) and photo-diode array detector. Samples were prepared by filtration with a 2 μm syringe filter. The mobile phase consisted of a 100% methanol solution (Solution A) and 1% acetic acid (Solution B) in nanopurewater. The method was as follows: 0.35 ml min^{-1} of 4% Solution A and 96% of Solution B for 4.5 min followed by ramping A to 50% (50% B) and maintained until min 6. The composition of the mobile phase was reverted to 4% Solution A and 96% Solution B and maintained for 8 additional min. The column was maintained at 45 °C and the sample reservoir at 15 °C. ACS grade *cis,cis*-MA was used to synthesize *cis,trans*-MA by heating in water at 75 °C for 25 min. *trans*-HDA (Sigma, St Louis, MO) and *cis,trans*-MA were used for UPLC calibration and as

references. Retention times of *cis, trans-MA* and *trans*-HDA are 6.4 min and 4.0 min and were analyzed at 295 nm and 231 nm respectively. Conversions and selectivities, were calculated with the following equations:

$$MA\,Conversion\,(\%) = 1 \quad \frac{[MA]_t}{[MA]_0} \quad *100$$

$$HDA\ Selectivity\ (\%) = \frac{[HDA]_t}{[MA]_0 \quad [MA]_t} *100$$

Voltage and Ph studies were performed on MA in K_2SO_4/H_2SO_4 electrolyte solutions in order to keep ionic strength and conductivity constant. A 0.1 M K_2SO_4 with 1.41 Mm MA solution was mixed in varying ratios with 0.1 M H_2SO_4 and 1.41 Mm MA. The solutions with the specified Ph were ECH in the 11 Ml small volume cell. In-between experiments, the electrodes were rinsed with DI water and the lead electrode was polished with a kimwipe.

ECH of the fermentation medium was performed as follows. Directly after the fermentation, 11 Ml of broth was put in a small volume cell. At the start of the ECH, the surface of a lead electrode was polished with a kimwipe and used as the electrocatalyst. In-between the successive batch reactions, the lead electrode was gently dipped in deionized water, dried, and put in another 11 Ml solution of the fermentation medium. In addition, the counter and reference electrodes were washed with deionized water in-between each run. To achieve a 94% yield of 3-hexenedioic acid (HDA) the Ph of the fermentation broth was decreased to 2.0 by dropwise adding 0.5 M H_2SO_4. Additionally, the concentration of MA was increased to 2.4 Mm.

6.1.2. Separations

Recovery of HDA was accomplished by concentrating ~2 L of cell-free fermentation broth containing the highest titer of HDA under basic conditions, followed by activated carbon filtration, and crystallization at reduced Ph and temperature. A yield of 67% and a high purity of 98% were obtained (**Supplementary Figure 3**).

Specifically, after removing cells from 2 L fermentation broth, the Ph of the cell-free broth was brought from 3 to 8.5 with 10 M NaOH to increase the solubility of HDA. The basic solution was run through a 0.2-μm filter to remove any solid impurities remained in the broth. Following filtration, the broth was concentrated in a vacuum evaporation set up, from ~2 L to 80 Ml and re-filtered through a 0.2-μm filter, separating the impurities that precipitated during concentration. The concentrate was then dripped through a 15-mm deep Norit CN1 activated carbon bed (preconditioned with 10 Mm NaOH), adsorbing soluble impurities including colorants and odor-causing compounds. The subsequent filtrate was acidified to Ph 1.5 using 18.4 M sulfuric acid and crystalized at 4 °C overnight. Crystals of HDA were recovered using vacuum filtration with a Whatman 50 filter, washed with 4 °C 10 Mm sulfuric acid, and placed in a desiccator to dryness.

6.1.3. Polymerization

The polycondensation reaction between *trans*-3-HDA and HMDA was adapted from the synthesis of nylon 6,6. Specifically, t*rans*-3-HDA (TCI America) was dissolved in methanol and mixed with a 1:1 mol ratio of HMDA dissolved in methanol. The resulting solution was heated in a round bottom flask at 60 °C. The liquid was decanted from the precipitated salt. The precipitated salt was subsequently washed with methanol, decanted, and left to dry in a fume hood. The solid was then mixed at a 0.86 mass ratio with deionized water. The resulting

solution was put in an aluminum weigh pan and heated at 7.5 °C min^{-1} to 250 °C in a tube furnace under flowing ultra-high purity nitrogen. The sample dwelled at the temperature for 30 min before cooling. It was not uncommon during the synthesis of UPA 6,6 to produce a slightly yellow colored polymer. The same synthesis procedure was applied to adipic acid and HMDA to form nylon 6,6.

6.2. Supplementary Tables

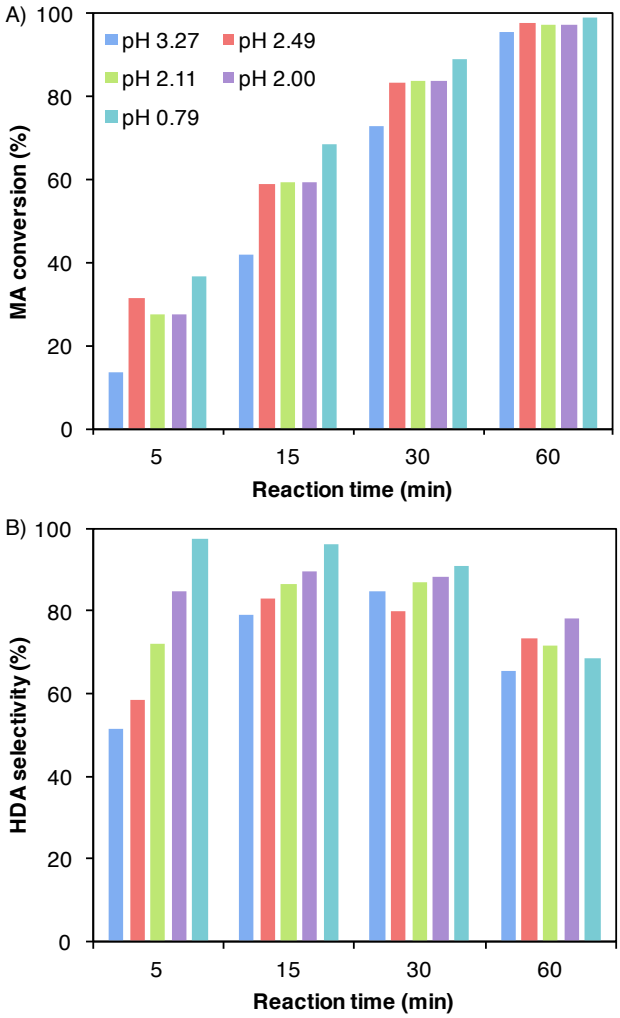
Supplementary Table 1: Primers used for gene knockout in S. cerevisiae InvSc1.

Fragment	**Primer**
aro1-Upstream	F: caacatattctcgatgtgct
	R: taagataattgtatattacg
aro1-Downstream	F: caataatatatctatccttt
	R: tacctgttcagtcgatacgt
pdc1-Upstream	F: tttcaatcattggagcaatc
	R: ataattagagattaaatcgc
pdc1-Downstream	F: tttgattgatttgactgtgt
	R: agtcagaagagcatacataa

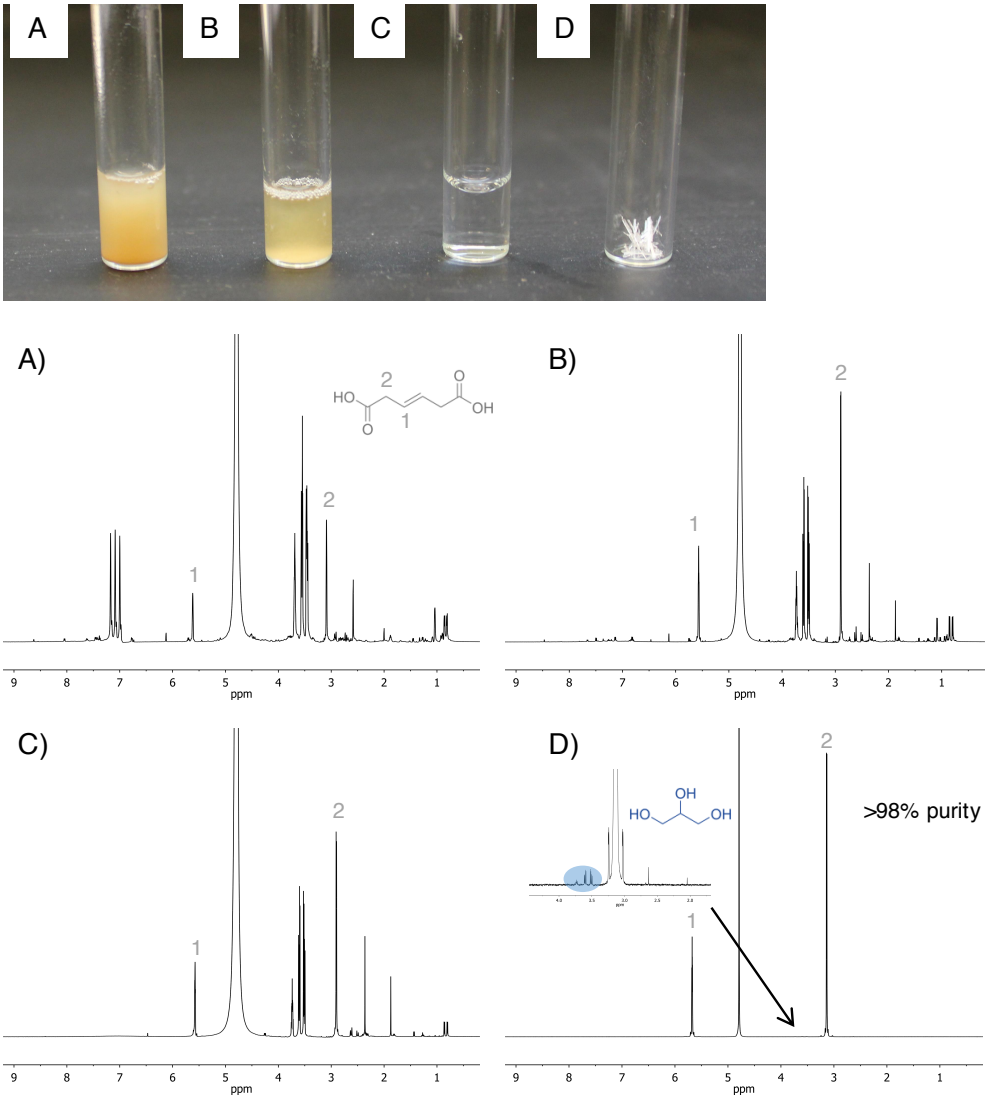
Supplementary Table 2: Triple detection SEC results of triplicate injections of UPA 6,6. Average recovery of all injection was 101.05 ± 0.72 %

Sample ID	Injection	V_{peak} (Ml)	M_n (Da)	M_w (Da)	M_z (Da)	M_w/M_n (PDI)	[η] (dl/g)	R_h (nm)	M-H α	M-H Log K
UPA 6,6	1	14.831	11,516	40,156	237,526	3.487	0.3142	5.124	0.525	-2.823
	2	14.826	10,255	39,336	349,171	3.836	0.3131	5.057	0.519	-2.794
	3	14.819	14,941	41,165	175,890	2.755	0.3097	5.24	0.551	-2.961
	Average	**14.825**	**12,237**	**40,219**	**254,196**	**3.359**	**0.3123**	**5.14**	**0.532**	**-2.859**
	σ	0.0049	1,980	748	71,717	0.45	0.0019	0.08	0.014	0.073
	% RSD	0.03%	16.18%	1.86%	28.21%	13.40%	0.61%	1.47%	2.61%	2.55%
Nylon 6,6	1	14.924	15,958	35,568	95,554	2.229	0.4538	5.831	0.573	-2.884
	2	14.968	19,547	37,217	95,573	1.904	0.4470	5.941	0.579	-3.006
	3	14.891	17,951	36,057	98,070	2.009	0.4592	5.900	0.580	-2.915
	Average	**14.928**	**17,819**	**36,281**	**96,399**	**2.047**	**0.4533**	**5.890**	**0.583**	**-2.935**
	σ	0.0315	1,468	692	1,182	0.135	0.0050	0.05	0.010	0.052
	% RSD	0.21%	8.24%	1.91%	1.23%	6.61%	1.10%	0.77%	1.73%	1.76%

6.3 Supplementary Figures



Supplementary Figure 1: ECH of model solutions of MA with varying Ph. The Ph in the solution was changed by adding various ratios of K_2SO_4 and H_2SO_4 to maintain the ionic conductivity. An ECH of the solution was run at -1.5 V for 1 h with the Pb catalyst. A) MA conversion and B) HDA selectivity.



Supplementary Figure 2: Photographs and 1H NMR spectra, 600 MHz, D2O during the separation of HDA following fermentation. The fermentation broth containing HDA was concentrated through distillation at Ph 8.5 (A). After concentration, the solid products were removed with a 0.2-µm filter (B) and the following solution was purified with activated carbon (C). Following the purification, HDA was recovered by crystallization at Ph 1.5 and 4°C overnight (D). Purity was obtained through liquid chromatography with a photodiode array detector

CHAPTER 4

INVESTIGATING STRAIN DEPENDENCY IN THE PRODUCTION OF AROMATIC COMPOUNDS IN *SACCHAROMYCES CEREVISIAE*

A manuscript published in the journal of Biotechnology and Bioengineering

Authors: Miguel Suástegui[1], Weihua Guo[2], Xueyang Feng[2,*], Zengyi Shao[1,*]

[1]Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

[2]Department of Biological Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

[*]Corresponding authors

Abstract

Although *Saccharomyces cerevisiae* is the most highly domesticated yeast, strain dependency in biotechnological processes still remains as a common, yet poorly understood phenomenon. To investigate this, the entrance to the aromatic amino acid biosynthetic pathway was compared in four commonly used *S. cerevisiae* laboratory strains. The strains were engineered to accumulate shikimate by overexpressing a mutant version of the pentafunctional ARO1 enzyme with disrupted activity in the shikimate kinase subunit. Carbon tracing and ^{13}C metabolic flux analysis combined with quantitative PCR, revealed that precursor availability and shikimate production were dramatically different in the four equally engineered strains, which were found to be correlated with the strains' capacity to deal with protein overexpression burden. By implementing a strain-dependent approach, the genetic platform was reformulated, leading to an increase in yield and titer in all strains. The highest producing strain, INVSc1-SA3, produced 358 mg L^{-1} of shikimate with a yield of 17.9 mg $g^{-1}_{glucose.}$ These results underline

the importance of strain selection in developing biological manufacturing processes, demonstrate the first case of high production of shikimate in yeast, and provide an appropriate platform for strain selection for future production of aromatic compounds.

1. Introduction

The metabolic pathways for the biosynthesis of aromatic compounds are present in many microorganisms and plants, and lead to the production of three key amino acids, *i.e.,* tryptophan, phenylalanine, and tyrosine. In certain plants such as *Papaver somniferum* (opium poppy) and *Catharanthus roseus*, these amino acids become intermediates in the production of more complex molecules that range from pigments, tannins, and alkaloids (Bohlmann et al., 1996; Facchini, 2001; Maeda and Dudareva, 2012). Given the versatile applications of these compounds and their great market opportunities (Evans, 2014), studying the relevant endogenous pathways has been central in microbial metabolic engineering efforts for the production of aromatic compounds. Some of the molecules that have reached considerable biotechnological production levels in *Escherichia coli* include shikimate (SA) (87 g/L) (Chandran et al., 2003), muconic acid (30 g/L) (Niu et al., 2002), and phenol (9.8 g/L) (Kim et al., 2014).

The production of more specialized aromatics such as flavonoids and alkaloids requires the functional expression of membrane-bound plant cytochrome P450 enzymes, which is feasible in yeast, but remains as a challenge in *E. coli* (Leonard et al., 2007). For instance, naringenin and resveratrol can only be converted through tyrosine in *E. coli*, but can be synthesized from both tyrosine and phenylalanine in *S. cerevisiae* due to the functional expression of cytochrome P450 reductase (Koopman et al., 2012; Shin et al., 2012), and the most recent breakthrough,

the biosynthesis of tyrosine-derived opioids (Galanie et al., 2015; Minami et al., 2008; Thodey et al., 2014), demonstrates the great potential of using yeast as a host to produce plant secondary metabolites. However, due to intrinsically complicated genetic and metabolic regulations, the efforts in engineering the central carbon metabolism to increase the production of aromatics in yeast seem somewhat ineffective. The reported yields in *S. cerevisiae* are far from reaching the theoretical values (Curran et al., 2012; Hansen et al., 2009; Koopman et al., 2012; Rodriguez et al., 2015), evidencing the demand for extensive strain development efforts.

A key step in the development of microbial platforms for the production of valuable chemicals is the appropriate selection of the host strain. Although it may sound trivial, it becomes more and more evident that selecting strains even among those belonging to the same species is key to ensuring optimal titers, yields, and productivities. Different *S. cerevisiae* strains, varying slightly in genetic makeups, unexpectedly display disparate behaviors (Strucko et al., 2015). The differences can be attributed to the varying capabilities of maintaining episomal plasmids (Karim et al., 2013), expressing proteins (Caesar et al., 2007), hosting heterologous pathways (Du et al., 2012), and responding to signaling processes (Kummel et al., 2010). However, the underlying basis for these varying phenotypes remains unclear and may vary case by case when engineering metabolic pathways in yeasts.

The purpose of this work was to analyze the metabolic differences among four commonly used *S. cerevisiae* strains (YSG50, BY4741, INVSc1, and BY4743) to understand the precursor dynamics that leads to high production of aromatic compounds. SA was chosen as the target compound since it is an early precursor for all the high-value aromatic compounds produced through this pathway, and itself is the starter compound of the commercial influenza antiviral Tamiflu® (Bochkov et al., 2012; McQuade and Blair, 2015). Although being equally

engineered, the four strains produced SA at dramatically different levels. ^{13}C-metabolic flux analysis (^{13}C-MFA) was implemented to systematically characterize the metabolic differences in the SA producing strains. The results demonstrated that the flux distribution in the central metabolism was very well conserved among all strains, but that the efficacy of producing SA was strongly dependent on the ability of each strain to auto-adjust the copy numbers of the episomal vectors. Studying the behavior of the engineered strains led to developing a new metabolic engineering rationale based on a low-copy plasmid, as a result enabling higher SA accumulation in all four strains. The work sheds light on increasing the titers of other aromatic amino acid-derived compounds in yeast.

2. Materials and Methods

2.1. Strains and media

Two haploid and two diploid strains of *S. cerevisiae* were analyzed in this study (**Table 1**). The haploid strains include YSG50 (*MATα ade2-1 ade3Δ22 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100)* and BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*); the diploid strains include INVSc1 (*MATa/α his3Δ1/his3Δ1 leu2/leu2 trp1-289/trp1-289 ura3-52/ura3-52 MAT*), and BY4743 (*MAT*a/α *his3*Δ*1/his3*Δ*1 leu2*Δ*0/leu2*Δ*0 LYS2/lys2*Δ*0 met15*Δ*0/MET15 ura3*Δ*0/ura3*Δ*0*). The wild type strains were propagated in YPAD media (1% yeast extract, 2% peptone, 0.01% adenine, and 2% dextrose). After plasmid transformation, yeast strains were cultured in Synthetic Complete media lacking uracil and histidine (0.5% ammonium sulfate, 0.16% yeast nitrogen without amino acid and ammonium sulfate, complete amino acid supplement mix lacking uracil and histidine, and 2% dextrose). The minimal growth medium composition used for labeling experiments was the same as reported previously with 20 g/L glucose as the sole carbon substrate (Feng and Zhao, 2013a; Feng and Zhao, 2013b). When

culturing yeast strains with ^{13}C-labeled glucose, the regular glucose was replaced with 80% [1-^{13}C] glucose and 20% [U-^{13}C] glucose (Cambridge Isotope Laboratories, Inc., Tewksbury, MA). All chemicals were purchased from Fisher Scientific (Pittsburgh, PA). *E. coli* BW25141 was cultured in Luria Bertani (LB) medium supplemented with 100 μg Ml^{-1} ampicillin.

2.2. Plasmid construction

All plasmids constructed in this study (**Table 1**) were derived from the PRS shuttle vector series (New England BioLabs, Ipswich, MA). To assemble the genetic cassettes, the DNA Assembler technique (Shao et al., 2012; Shao and Zhao, 2013) was implemented. All primers in this work were synthesized by IDT (Coralville, IA). The aspartic acid at the position of 920 in the wild type ARO1 was substituted to alanine by incorporating the mutation into PCR primers (*fwd*:gttacaaattagttgaccta***gcg***gagctgtttgagcaaca, *rev*: tgttgctcaaacagctc***cgc***taggtcaactaatttgtaac) and the variant was cloned downstream of the *S. cerevisiae* GPD constitutive promoter and upstream of the PYK1 terminator. The expression cassette was assembled in the PRS426 backbone digested with *Xho*I, yielding the plasmid pRS426-*aro1*$_{D920A}$ (**Supplementary Figure 1**). Plasmid pRS413-*aro4*$_{K229L}$-*tkl1* was constructed in the previous study (Suastegui et al., 2016) and used here to enhance the entrance of carbon into the SA pathway. Plasmid pRS413-*aro1*$_{D920A}$-*aro4*$_{K229L}$-*tkl1* was constructed by assembling the *Xho*I digested pRS413-*aro4*$_{K229L}$-*tkl1* with the PCR-amplified *aro1*$_{D920A}$ cassette with the ends overlapping to the pRS413 backbone and TPI1 promoter. Similarly, to construct pRS426-*aro1*$_{D920A}$-*aro4*$_{K229L}$-*tkl1*, plasmid pRS426-*aro1*$_{D920A}$ was digested with *BamH*I and re-assembled by addition of the PCR cassette *aro4*$_{K229L}$-*tkl1* with the overlapping ends to the PYK terminator and the backbone of pRS426. Plasmid transformation was done by the lithium acetate technique (Kawai et al., 2010)

Table 1: Plasmids and strains constructed in this work.

Plasmid	**Expression Cassette**	**Features**
pRS413-*aro4*$_{K229L}$-*TKL1*	TPI1p-*scaro4*$_{K229L}$-TDH2t ADH1p-*scTKL1*-ADH1t	Feedback insensitive DHAP synthase. Transketolase 1 to increase E4P pool.
pRS426-*aro1*$_{D920A}$	GPDp-*scaro1*$_{D920A}$-PYK1t	Mutant version of pentafunctional ScARO1 with disrupted activity of SA kinase subunit.
pRS426-*aro1*$_{D920A}$ – *ecaroL*	GPDp-*scaro1*$_{D920A}$-PYK1t TEFp-*ecaroL*-HXT7t	*E. coli aroL* to complement the activity of mutant *aro1*$_{D920A}$
pRS413-*aro1*$_{D920A}$ *aro4*$_{K229L}$-*TKL1*	GPDp-*scaro1*$_{D920A}$-PYK1t TPI1p-scARO4$_{K229L}$-TDH2t ADH1p-ScTKL1-ADH1t	Complete pathway in pRS413 backbone for accumulation of SA.
pRS426-*aro1*$_{D920A}$ *aro4*$_{K229L}$-*TKL1*		Complete pathway in pRS426 backbone for accumulation of SA.
Strains	**Genotype / Plasmid**	
YSG50	*MATα; ade2-1; ade3 Δ22; ura3-1; his3-11,15; trp1-1; leu2-3,112; can1-100*	
INVSc1	*MATa/α; his3 Δ 1/his3 Δ1; leu2/leu2; trp1-289/ trp1-289; ura3-52/ ura-52 MAT*	
BY4741	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*	
BY4741 *Δaro1*	BY4741 *aro1*::*KanMX*	
BY4743	*MAT*a/α; *his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; LYS2/lys2Δ0; met15Δ0/MET15; ura3Δ0/ura3Δ0*	
SA0	pRS426 & pRS413	
SA1	pRS426-*aro1*$_{D920A}$	
SA2	pRS426-*aro1*$_{D920A}$ & pRS413 *aro4*$_{K229L}$-*TKL1*	
SA3	pRS426 & pRS413-*aro1*$_{D920A}$-*aro4*$_{K229L}$-*TKL1*	
SA4	pRS426-*aro1*$_{D920A}$- *aro4*$_{K229L}$-*TKL1* & pRS413	

2.3. Biomass and metabolite analysis

Biomass growth in minimal media was monitored by optical density changes with an absorbance reader at 600 nm (OD_{600}) (Biotek Synergy 2 Multi-Mode Microplate Reader). Samples were taken every 2 to 3 hours until the late exponential phase was reached. Cell-free samples were used to measure extracellular metabolites by Shimadzu HPLC (Columbia, MD) equipped with an Aminex HPX-87H column (Bio-Rad, Hercules, CA) and Shimadzu RID-10A refractive index detector. The column was kept at 50°C, and 5 Mm sulfuric acid solution was used as the mobile phase with a constant flow rate of 0.6 Ml/min. Standard samples of ethanol, glucose, glycerol, and SA were purchased from ACROS Organics™ (Geel, Belgium).

2.4. Isotopomer analysis of proteinogenic amino acids and SA

The protocol for sample preparation and amino acid analysis was followed based on Feng *et al.* (Feng et al., 2012). During the late exponential phase, 1 Ml samples were collected from the cultures grown with the labeled glucose for proteinogenic amino acid profiling. The samples were centrifuged and the biomass was hydrolyzed with 6 M HCl at 100°C for 20 h, followed by drying and derivatization with 25 μL of tetra-hydrofuran and 25 μl of tert-butyldimethylsilyl ether at 70°C for 1h. To analyze the labeled SA, 1 Ml of the supernatant from the culture samples was dried overnight followed by derivatization with 60 μl of N,O-bis(trimethylsilyl)trifluoroacetamide at 70°C for 1 h. The derivatized samples were analyzed by gas chromatograph (GC2010, Shimadzu) equipped with a SH-Rxi-5Sil column (Shimadzu) and a mass spectrometer (QP2010, Shimadzu). Two types of charged fragments were detected by GC-MS for various amino acids (**Supplementary Table 1**): the $[M-57]^+$ group (containing unfragmented amino acids), and the $[M-159]^+$ or the $[M-85]^+$ group (containing amino acids that had lost an α-carboxyl group). The whole derivatized SA fragment was detected from the

sample supernatant. The mass distribution vectors (MDVs) were represented by M0, M1, M2, etc., which were the corresponding fractions of non-labeled, singly labeled, doubly labeled amino acids. The effects of natural isotopes on isotopomer labeling patterns were corrected by the previously reported algorithms (Wahl et al., 2004).

2.5. Metabolic Flux Analysis

MDVs were used to calculate the summed fractional labeling (SFL) values which were directly used in the MFA (Biomet Toolbox 2.0 (Garcia-Albornoz et al., 2014), based on MATLAB (MathWorks, Inc. MA). The central carbon metabolic model for the MFA was developed based on the KEGG database (http://www.genome.jp/kegg/), and it included the glycolysis pathway, pentose phosphate pathway, anaplerotic pathways, the tricarboxylic acid (TCA) cycle, the SA synthesis pathway, and the transport reactions between different cell compartments (**Supplementary Table 2**). For our metabolic flux analysis, flux estimation was repeated at least 50 times starting with different initial values generated by a genetic algorithm for all the fluxes to find a likely global solution. A fit of the simulated and measured SFLs was determined to be a global solution only after the solution fit was obtained at least twice using this method (**Supplementary Figure 2**).

2.6. Plasmid copy number assay

The copy number of the two plasmids used in this work was quantified using quantitative PCR (Qpcr). Cell cultures were grown in minimal media for 24 hours and cell amounts equivalent to 2 OD_{600} units were collected for total DNA extraction as previously described (Moriya et al., 2006). The primers were designed using the online PrimeTime© tool for Qpcr assays through IDT. The primer pair p-ura3 (*fwd*: tagcagaatgggcagacattac; *rev*: ggcctctaggttcctttgttac) was used to quantify the plasmid pRS426; the primer pair p-his3 (*fwd*: gtccacacaggtatagggtttc;

rev: gtgatggtcgtctatgtgtaagt) was used to quantify the plasmid pRS413. Plasmid copy number was normalized to total DNA with *alg9* as the reference gene (*fwd*: ccgttgccatgttgttgtatg; *rev*: gccaggaaattgtacgctaaac). A standard curve for each target was prepared using pure plasmid or genomic DNA samples.

3. Results and Discussion

3.1 Building the first yeast platform for SA production

Unlike prokaryotes and higher eukaryotes, the five steps involved in SA biosynthesis in yeast are catalyzed by a single pentafunctional enzyme ARO1 (Braus, 1991) (**Figure 1**). Knocking out individual steps of this pathway without affecting the rest of the catalytic functions in yeast could be a cumbersome task. Hence, a single amino acid substitution in the active site of the SA kinase subunit (ARO1B) is preferred to halt the conversion of SA to shikimate-3-phosphate (S3P). Initially, an amino acid sequence alignment with SA kinases from several prokaryotic and eukaryotic organisms was performed. Despite the genetic variations within these species, it was observed that the catalytic sites of the shikimate kinases (D920-*aro1* in *S. cerevisiae*) were 100% conserved (**Figure 2**). This residue was substituted by alanine, yielding the mutant enzyme $ARO1_{D920A}$ that was further cloned into plasmid pRS426 (pRS426-$aro1_{D920A}$) and transformed into the four *S. cerevisiae* strains (YSG50, INVSc1, BY4741, and BY4743). However, no SA was detected in the fermentation media in any of the strains (**Figure 3**). It was not a surprise since the endogenous synthesis of aromatic amino acids could have (i) channeled the accumulated SA into the downstream reactions, and (ii) feedback inhibited the

first step of aromatic amino acid biosynthesis catalyzed by 3-dehydroxy-D-arabino-heptulose-7-phosphate (DAHP) synthase.

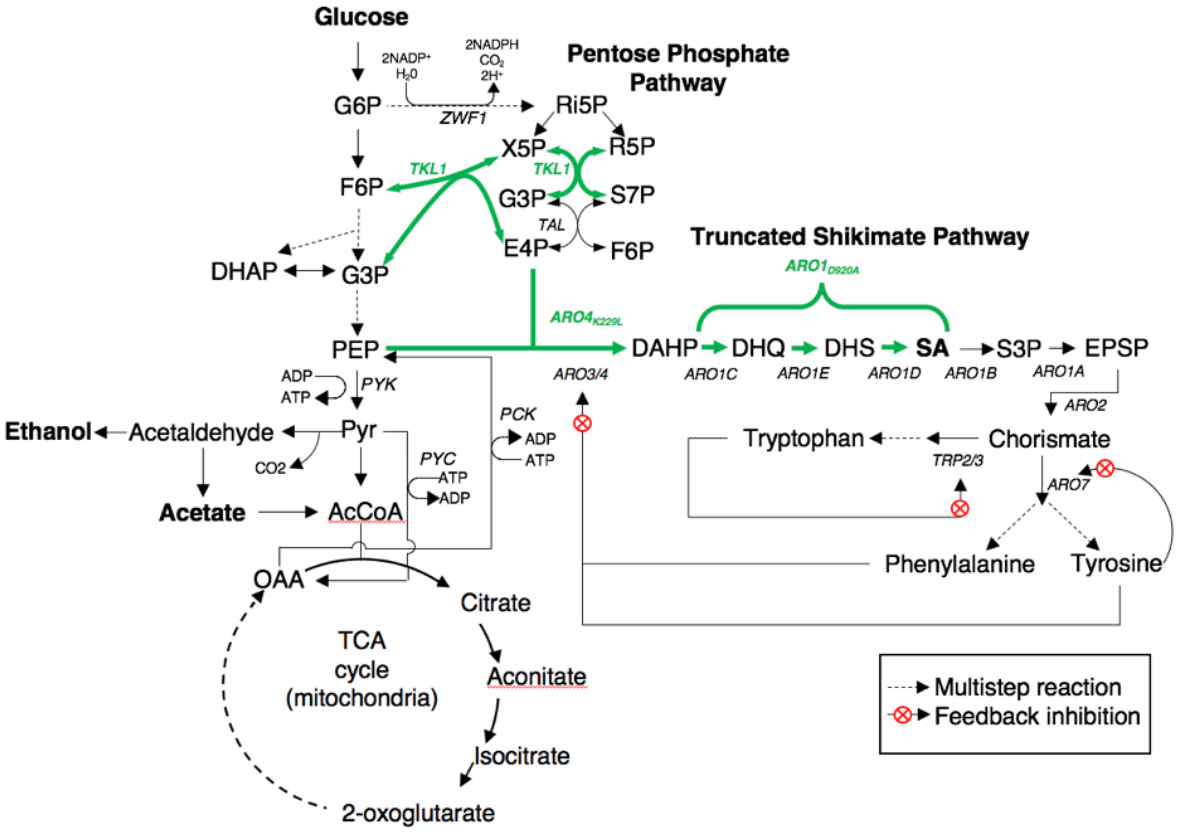


Figure 1: Pathway depicting engineering efforts for the accumulation of SA. The arrows in green represent the overexpressed enzymes for unlocking the production of SA. **Metabolites**: G6P, glucose-6-phosphate; G1P, glucose-1-phosphate; F6P, fructose-6-phosphate; G3P, glyceraldehyde-3-phosphate; DHAP, dihydroxyacetone phosphate; PEP, phosphoenolpyruvate; Pyr, pyruvate; AcCoA, acetyl-CoA; OAA, oxaloacetic acid; Ri5P, ribulose-5-phosphate; X5P, xylulose-5-phosphate; R5P, ribose-5-phosphate; S7P, sedoheptulose-7-phosphate; E4P, erythrose-4-phosphate; DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate; DHQ, dehydroquinoate; DHS, dehydroshikimate; S3P, shikimate-3-phosphate; EPSP, 5-enolpyruvyl-3-shikimate phosphate. Abbreviation of genes: *ZWF1,* glucose-6-phosphate 1-dehydrogenase; *PYK1*, pyruvate kinase; *PCK*, phosphoenolpyruvate carboxykinase; *PYC*, pyruvate carboxylase. Abbreviation of enzymes: TKL1, transketolase; ARO3/4, DAHP synthase isoenzymes; $ARO4_{K229L}$, feedback insensitive DAHP synthase; ARO1, pentafunctional enzyme converting DAHP to EPSP; subunits – C: DHQ synthase, E: DHQ dehydratase, D: DHS dehydrogenase, B: SA kinase, A: ESPS synthase; $ARO1_{D920A}$, mutant ARO1 with disrupted activity of shikimate kinase subunit; TRP2, anthranilate synthase; ARO2, chorismate synthase; TRP3, indole-3-glycerol-phosphate synthase; ARO7, chorismate mutase

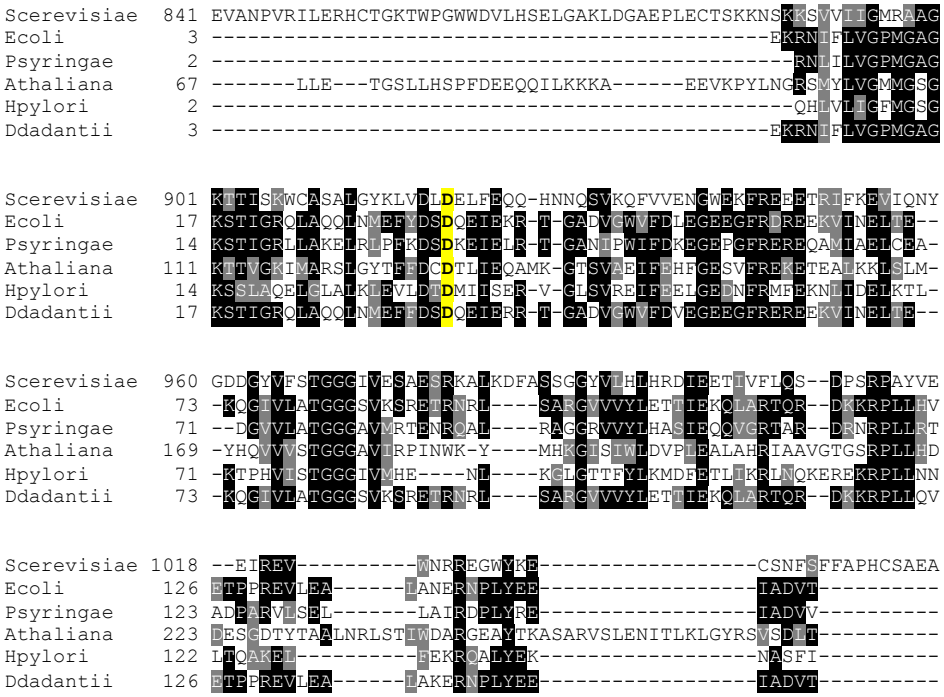


Figure 2: Sequence alignment of shikimate kinases from different species. The aspartic acid residue (D920 in *S. cerevisiae*) is conserved in the five selected species, from which the corresponding residue has been reported to be the catalytic residue. UNIPROT 92ccession number: *S. cerevisiae: aro1*B subunit (P08566), *Escherichia coli* (P0A6D7), *Pseudomonas syringae* (Q87V14), *Arabidopsis thaliana* (Q8GY88), *Helicobacter pylori* (P56073), and *Dickeya dadantii* (E0SJR7).

To characterize the activity of ARO1$_{D920A}$, a complementation experiment was carried out. The plasmid pRS426-*aro1*$_{D920A}$ was transformed into the strain BY4741 Δ*aro1* (GE Healthcare Bio-Sciences, USA) and plated onto medium lacking uracil and the three aromatic amino acids. Cell were incubated for two days and as expected no growth was observed, indicating that the novel mutant ARO1$_{D920A}$ is incapable of complementing growth in the Δ*aro1* background (**Figure 4a**). To confirm that the growth defect is caused exclusively by the disrupted kinase subunit (ARO1B) in the novel mutant ARO1$_{D920A}$, a new plasmid was constructed harboring the *E. coli* shikimate kinase gene *aroL* (pRS426-*aro1*$_{D920A}$ – *ecaroL)*. The strain BY4741 Δ*aro1* was transformed with the latter plasmid and after two days of culturing in media lacking uracil and aromatic amino acids, colonies were observed (**Figure 4b**). Complementation of growth by the *ecaroL,* was a definitive indication that the novel enzyme ARO1$_{D920A}$ remains functional in all of its subunits except for the subunit ARO1B, which catalyzes the conversion of shikimic acid to shikimate-3-phosphate.

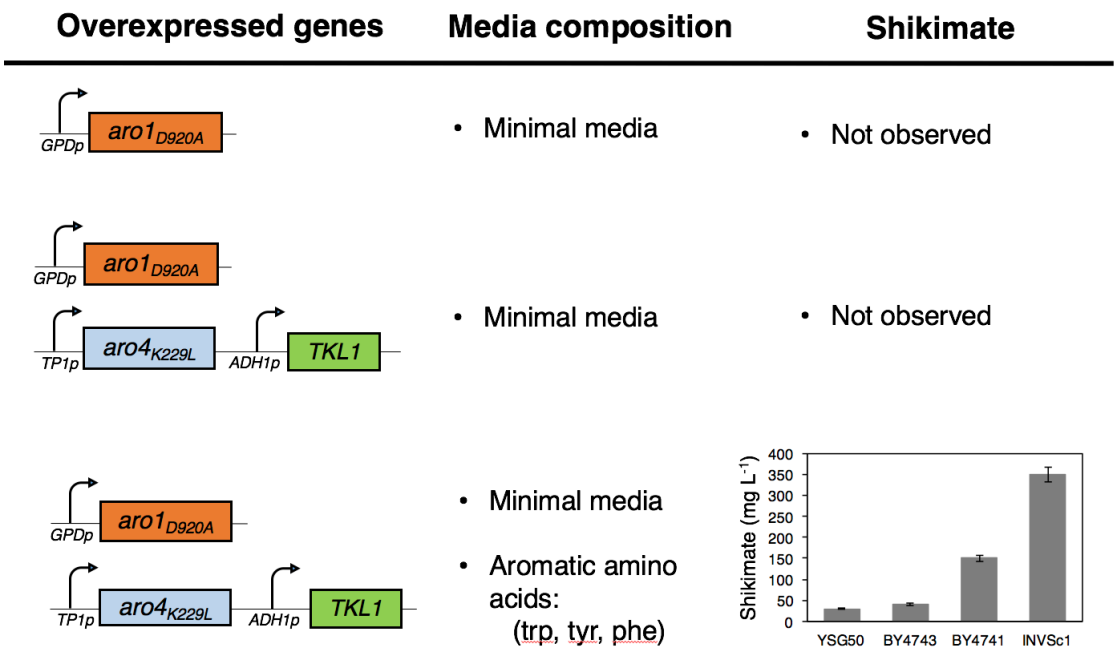


Figure 3: Media optimization for the production of SA. To allow the accumulation of SA in all fours strains, it was necessary to incorporate the three aromatic amino acids into the growth media, through which the conversion of SA into downstream metabolites was at least partially reduced and hence the positive outcome of expressing the mutant $ARO1_{D920A}$ was observed. Samples were taken after 24 h of fermentation. The variation is represented as the standard deviation from three biological replicates.

The plasmid pRS413-*aro4*$_{K229L}$-*tkl1* was transformed into the strains (Group SA2) which carries the tyrosine-insensitive DAHP synthase to overcome the feedback inhibition, as well as the transketolase gene *tkl1* to increase the flux of the precursor erythrose-4-phosphte (E4P) (Curran et al., 2012). These approaches have been successfully implemented in *E. coli* to produce SA and in *S. cerevisiae* for the production of dehydroshikimate-derived muconic acid (Chen et al., 2012; Cui et al., 2014; Curran et al., 2012; Niu et al., 2002). Nevertheless, SA was still not observable in any of the strains. We hypothesized that the genome-copy of *aro1* was causing a rapid channeling of SA into the production of aromatic amino acids. Hence, to completely shut down the native pathways downstream of SA, tryptophan, phenylalanine, and tyrosine were added to the minimal medium at low concentrations of 50 mg L^{-1}. The presence of these three amino acids blocked the endogenous flux into the downstream branches, and SA was finally observed (**Figure 3**). This result demonstrated that despite genotypic differences, all four *S. cerevisiae* strains had undergone strong regulatory controls on the aromatic amino acid biosynthesis. It is known that the presence of external aromatic amino acids inhibits the first committed step in the biosynthesis of aromatics (ARO3 and ARO4*)* and the downstream branch points (ARO7 and TRP2/3, **Figure 1**) (Braus, 1991). However, inhibition may also have occurred at the step converting SA to S3P (ARO1B), most likely by the reduced transcription rates of the wild type *aro1*, thus allowing SA to accumulate in the strains overexpressing the mutant ARO1$_{D920A}$. To our knowledge, this is the first report of success in engineering yeasts to accumulate a significant amount of SA.

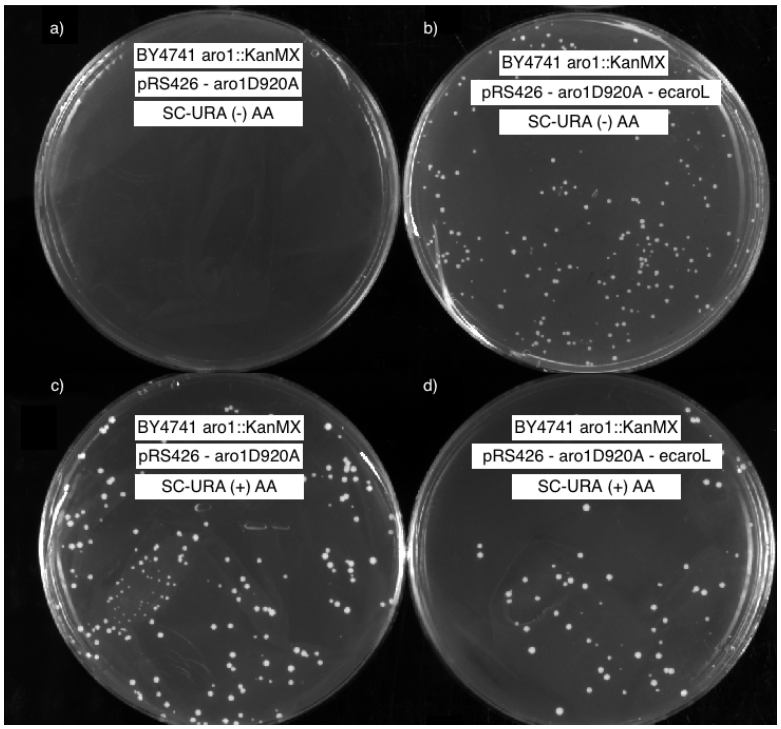


Figure 4: **Characterization of the novel mutant enzyme ARO1$_{D920A}$.** A) The strain BY4741 Δ*aro1* harboring plasmid pRS426-*aro1*$_{D920A}$ cannot grow in medium lacking aromatic amino acids (AA). B) Complementation of growth was achieved by constructing the plasmids pRS426-*aro1*$_{D920A}$ – *ecaroL,* which harbors the *E. coli* shikimate kinase gene *aroL.* C-D) Control plates containing aromatic amino acids.

3.2. Comparison of fermentation profiles

Although all strains were able to produce SA, the titers among the strains varied drastically; the lowest being 30 mg L^{-1} in strain YSG50 and the highest, 350 mg L^{-1} in strain INVSc1. Small batch fermentations in minimal media were carried out to investigate physiological characteristics. The consumption rates of glucose and the formation rates of SA, ethanol, and glycerol among the four strains during exponential phase were compared (**Table 2**, **Figure 5**). From the fermentation profiles it was easy to arrange the strains into two groups; the strain pair of BY4741 and BY4743, and the pair of YSG50 and INVSc1. Although all four strains had similar specific growth rates (0.13-0.17 h^{-1}), the BY4741 and BY4743 strains did have about 2-3 fold higher rates in terms of glucose consumption and ethanol production. The rate of glycerol production varied slightly, although it was not observable in strain YSG50. The production of SA revealed the highest difference among the strains. Apparently, no correlation between strain ploidy and SA production could be drawn since strains YSG50 and BY4743 (haploid and diploid, respectively) showed the lowest productions, with yields around 0.83 and 0.32 mmol $Gdcw^{-1}$, respectively. The strain BY4741 had a mid SA production with a yield of 2.38 mmol $Gdcw^{-1}$, and finally, strain INVSc1 had the highest yield of 8.1 mmol $Gdcw^{-1}$, 3.4-fold higher than strain BY4741, 9.8-fold higher than strain YSG50, and 25-fold higher than BY4743. These results suggested that, despite all the strains being equally engineered, either the expression of the pathway enzymes, the entrance into the aromatic amino acid pathway or both were strain dependent.

Table 2: Physiological characterization of SA producing strains. μ (h^{-1}), specific growth rate; r (mmol $Gdcw^{-1}$ h^{-1}), rate of consumption or production; Y_{max} (mmol $Gdcw^{-1}$), molar yield based on consumed glucose; titer (mg L^{-1}).

Strain		Glucose	Ethanol		Glycerol		SA		
	μ	r	r	Y_{max}	r	Y_{max}	r	Y_{max}	titer
YSG50-SA2	0.13	13.2	15.1	241.8	-	-	0.17	0.83	30
INVSc1-SA2	0.17	13.8	12.8	206.9	1.49	28.6	1.53	8.1	350
BY4741-SA2	0.15	28.7	34.8	246.4	4.79	36.1	1.36	2.38	150
BY4743-SA2	0.16	33.4	40.1	310.6	1.18	18.6	0.15	0.32	40

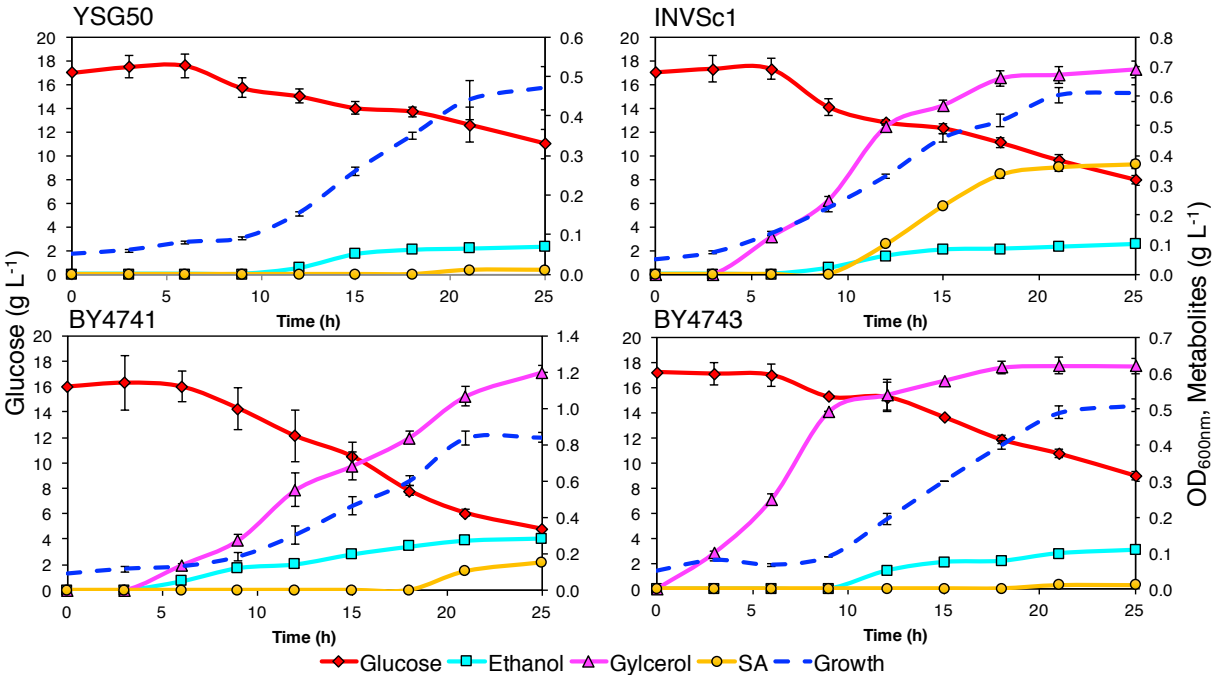


Figure 5: Fermentation profiles of strain groups SA2

3.3. Inspection of ^{13}C isotopic "finger prints" of aromatic amino acids

The most significant difference between strains in the group SA2 arose from comparing the ^{13}C isotopic "fingerprints" of the aromatic amino acids, namely phenylalanine and tyrosine Interestingly, the strains with the highest production of SA (INVSc1 and BY4741) showed the highest percentage of labeled phenylalanine and tyrosine (m1:m9), whereas the strains with the lowest production (YSG50 and BY4743) showed a lower labeling distribution (**Figure 6**). This was a clear indication that not only were the strains INVSc1 and BY4741 able to accumulate SA in higher titers, but they also had a higher pull of carbon from the downstream metabolic branch leading to the production of phenylalanine and tyrosine. It is crucial to mention that these observations were not due to different extracellular amino acid uptake, since measurement of the fermentation supernatant showed equal uptake rates (data not shown here). Analyzing the isotopic labeling patterns of leucine, another non-labeled amino acid added in the culture medium for all four strains also showed similar labeling patterns; the non-labeled leucine (m0) constituted more than 93% in all strains. This confirmed that the high producing SA strains had more active aromatic amino acid biosynthetic pathway. It is noteworthy that although the endogenous entrance of carbon into the aromatic pathway was supposedly blocked by the supplementation of aromatic amino acids, *de novo* production of tyrosine and phenylalanine was observed in all strains. Therefore, to maximize SA accumulation, future work should target the implementation of a more stringent blockage of the downstream branches.

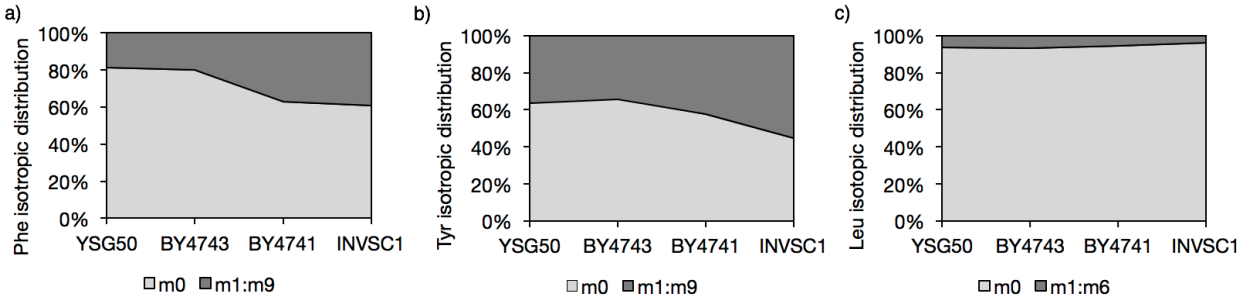


Figure 6: Isotropic distribution of labeled and non-labeled amino acids. Strains INVSC1-SA2 and BY4741-SA2 showed higher percentages of labeled A) phenylalanine and B) tyrosine compared to YSG50-SA2 and BY4743-SA2. This corroborated the higher fluxes into the biosynthesis of aromatic compounds in the strains with higher SA levels. C) Analysis of leucine, another non-labeled amino acid, showed equal distribution across all four strains.

3.4. Comparison of flux distributions in key nodes

To further investigate the metabolic differences in a quantitative manner, the isotopomer labeling patterns, *i.e.*, mass distribution vectors, of proteinogenic amino acids were utilized to obtain the metabolic fluxes through the central carbon metabolism of the *S. cerevisiae* strains. Similar fluxes were found in the glycolytic pathway up to the formation of phosphoenolpyruvate (PEP), as well as in the oxidative pentose phosphate pathway (PPP) (**Figure 7**). The main variation resulted from the flux entrance into the SA synthesis pathway (*v11*). This reflected difference in the availability of the precursors PEP and E4P to enter the pathway. Less than 8% of the flux into PEP (*v5),* ranging from 143 to 172, was channeled into the formation of SA in all the strains.

Therefore, a potential strategy to increase the availability of PEP is to lower the activity of the pyruvate kinase encoded by *PYK1* (**Figure 1**). This strategy has been implemented in the bacterium *Bacillus subtilis* through (i) deleting *PYK1*, resulting in 8% increase of SA (Liu et al., 2014), and (ii) reducing the activity by means of a mutant *PYK1* (Fry et al., 2000) or by

replacing its promoter to an inducible one which resulted in the least detrimental growth (Zhu et al., 2005). A similar approach could be implemented in *S. cerevisiae*, since complete removal of *PYK1* would impede the cells from growing in fermentable sugars (Sprague, 1977).

Besides the difference in the PEP supply, another major difference occurred in the reductive PPP, which is controlled by the Tkl1 and the transaldolase (Tal1) enzymes, catalyzing the transfer of C2 and C3 units from a ketose donor to an aldose acceptor. The enzyme Tkl1 has been shown to be indispensable in the production of aromatic amino acids; deletion of this enzyme in *S. cerevisiae* made the strain auxotrophic to aromatic amino acids (Schaaff-Gerstenschlager et al., 1993). Furthermore, *TKL1* has been overexpressed to increase the production of DHS-derived muconic acid in combination with the deletion of the gene *ZWF1* to force carbon to enter PPP at the non-oxidative node (**Figure 1**). However, only a small increase in muconic acid was observed (from ~ 30 mg L^{-1} to 40 mg L^{-1}), which could be correlated with the preference of TKL1 to catalyze the opposite reaction (Curran et al., 2012). Although this strategy has been successfully implemented in *E. coli* (Niu et al., 2002; Patnaik and Liao, 1994), a recent publication studying the production of muconic acid in a co-culture system showed that this approach overexpression resulted in a lower titer (Zhang et al., 2015). Expectedly, the flux through the reaction *v10* evidently differed among the strains; the lowest value, 1.1, corresponded to the highest producer, while the lowest producer showed a flux value of 17.5. This confirmed that INVSc1 had a stronger pull of carbon into the aromatic amino acid pathway, whereas in YSG50 and BY4743 more carbon returned to glycolysis at the node of E4P. An important point to highlight is that between the two precursors for aromatic compounds, E4P was clearly the limiting one. For instance, in strain INVSc1, despite that about 90% of the flux of E4P ended up entering the SA pathway (*v11/v9)*, it was only

equivalent to 8% of the calculated flux of PEP (*v11*/*v5)*. Thus, engineering the node to increase the availability of E4P is key to maximizing the production of aromatic compounds. This could entail pulling more carbon into the non-oxidative PPP by means of a mutant TKL incapable of converting E4P back to F6P (Nilsson et al., 1997), or by implementing global transcriptional manipulations to increase the transcription rates of multiple genes in this node coupled with a fast-screening method (Iraqui et al., 1998; Kim et al., 2015; Lee and Hahn, 2013).

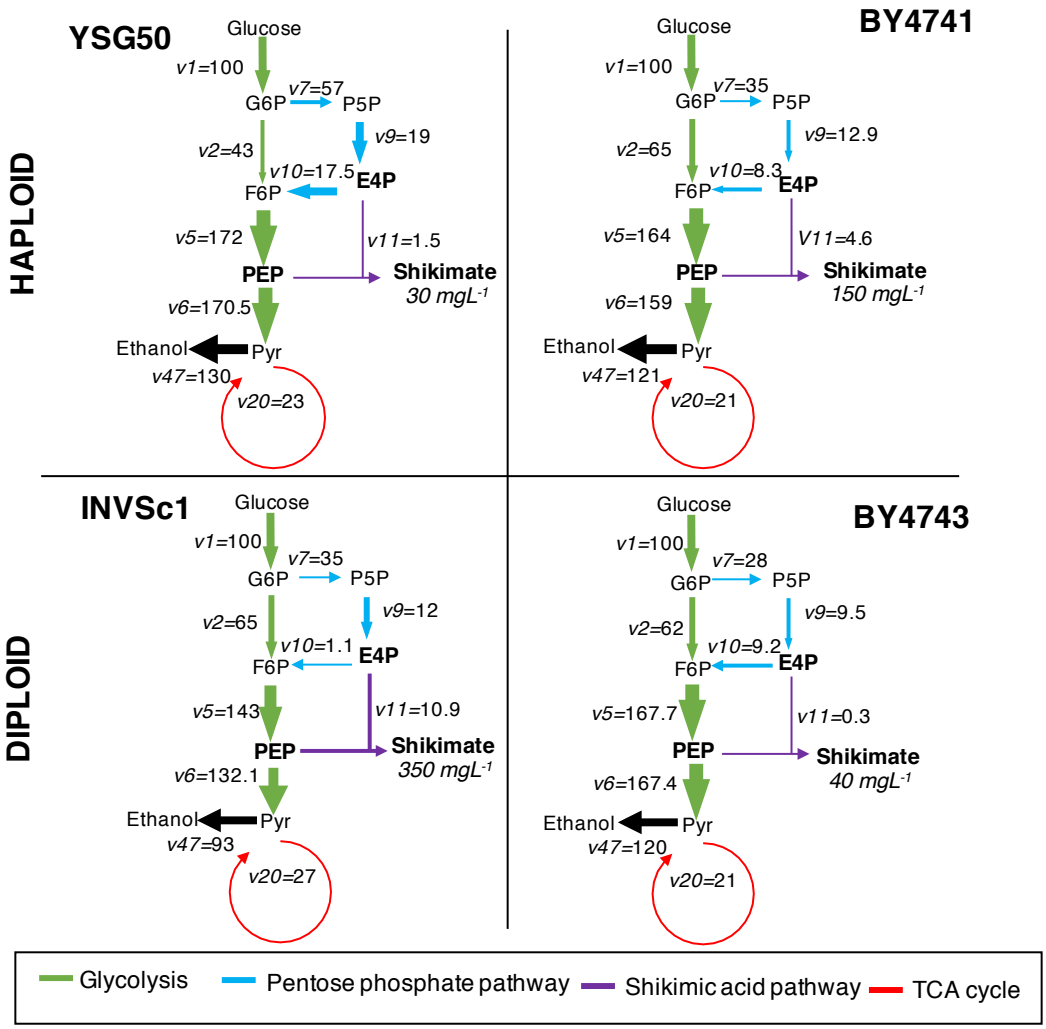


Figure 7: Major intracellular metabolic flux distributions in the engineered strains in group SA2 (**Table 1**). All fermentations were performed in batch mode supplemented with 20 g L^{-1} of 80% [1-^{13}C] glucose and 20% [U-^{13}C] glucose. The fluxes are expressed as normalized values relative to specific glucose uptake rate (**Table 2**). These values represent a global solution obtained *via* isotopomer modeling of the ^{13}C labeled proteinogenic amino acids. To emphasize the flux differences in the SA pathway, the thickness of the arrows of reactions *v9* to *v11* was normalized by a value of 5; a value of 25 was used otherwise. P5P: pentose phosphate sugar.

3.5. Pathway balancing for enhanced production of SA

Carbon tracing and ^{13}C-MFA corroborated that the four equally engineered strains had dramatically different flux distributions through the SA pathway stretching from the upstream precursor synthesis to the downstream aromatic amino acid production, which was likely caused by the differential expression of the three cloned enzymes. Initially, we hypothesized that the plasmid copy numbers could vary significantly from one strain to another, causing the major metabolic differences observed, despite all strains being equally engineered. Previous research has shown that specific genetic elements on the plasmid such as the selection marker, as well as the segregation element (centromere or 2μ) play an important role in the yeast's fitness with some correlations to the strain's ploidy (Karim et al., 2013). Since our platform consisted of plasmids harboring structural genes under the transcriptional control of strong constitutive promoters, we further reasoned that the engineered strains might have had very different capacities to deal with the energy burden posed by plasmid synthesis and enzyme overexpression. This effect was first recognized based on the observation that the optical densities at the late exponential phase of strain cultures in the group SA2 were reduced compared to the strains carrying the empty plasmids (SA0 strains, **Figure 8a**), but interestingly the growth reduction did not correlate with SA production level. This observation excluded SA toxicity as the major factor causing the growth reduction, suggesting that the actual copy numbers in the four strains might not reflect the behaviors of typical 2μ and centromere as the segregation elements.

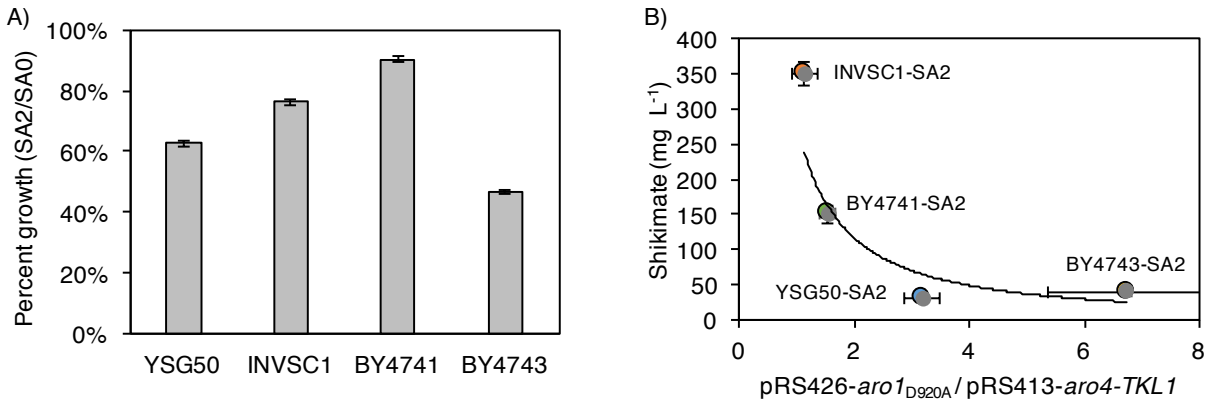


Figure 8: Phenotypic comparison of *S. cerevisiae* engineered for the production of SA. A) Growth percentage of strains in the group SA2 versus strains carrying empty plasmids (group SA0, **Table 1**). B) Correlation between SA production and plasmid maintenance. The ratio between plasmid pRS426-*aro1*$_{D920A}$ and pRS413-*aro4*$_{K229L}$-*TKL1* was calculated and plotted on the x-axis. Samples were taken after 24 h of growth in minimal media. Variation is represented as the standard deviation from three biological replicates.

Measurement of the plasmid copy numbers revealed interesting insights on the capacity of each strain to cope with the stress caused by excessive enzyme expression. It appeared that plasmid pRS426-*aro1*$_{D920A}$ had a significant impact on the cell's fitness, and most importantly, on the production of SA. As shown in **Figure 8b**, the copy number ratio between the plasmid pRS426-*aro1*$_{D920A}$ and pRS413-*aro4*$_{K229L}$-*tkl1* was calculated and a clear negative correlation was identified. The strains with the highest production of SA (INVSc1 and BY4741) had a plasmid ratio closer to 1, whereas the poor-producing strains (YSG50 and BY4743) had a ratio between 3 and 7. It was also noticed that although the plasmid pRS426-*aro1*$_{D920A}$ had the 2μ segregation element, the copy number was dramatically reduced to resemble a centromere-like plasmid. The absolute copy numbers of the plasmids were compared between the strain groups SA0 and SA2 (**Figure 9**). The reduction in plasmid copy number was evident in all strains in the group SA2; the most significant reduction occurred in the plasmid pRS426-*aro1*$_{D920A}$ in the strain INVSc1-SA2, where the copy number was reduced 95% compared to

the empty plasmid. Contrary to our findings, Karim et al., (2013) showed that using histidine and uracil as selection markers result in similar growth rates and plasmid copy number when comparing strains BY4741 and BY4743. In our study, we did observe major discrepancies that can be attained to 1) co-expression of two distinct episomal plasmids and 2) constitutive overexpression of structural genes. Although the four, wildtype strains studied here have slight differences in their auxotrophic markers, no correlation with their capacity to accumulate shikimate can be observed. For instance, the only major auxotrophic difference between strains YSG50 and INVSc1 is the deletion of *ade1* in the former, yet these strains have the most evident disparity in accumulation of shikimate. Similarly, supplementation of L-trp in the growth medium brings the strains INVSc1 and BY4743 to the same auxotrophic level, but their capacity to produce shikimic is considerably different. We hypothesized that overexpression of *aro1*$_{D920A}$ in the pRS426 background had a negative impact on SA production. Hence, only the strains capable of adjusting this plasmid to a very low copy number could produce higher amounts of SA.

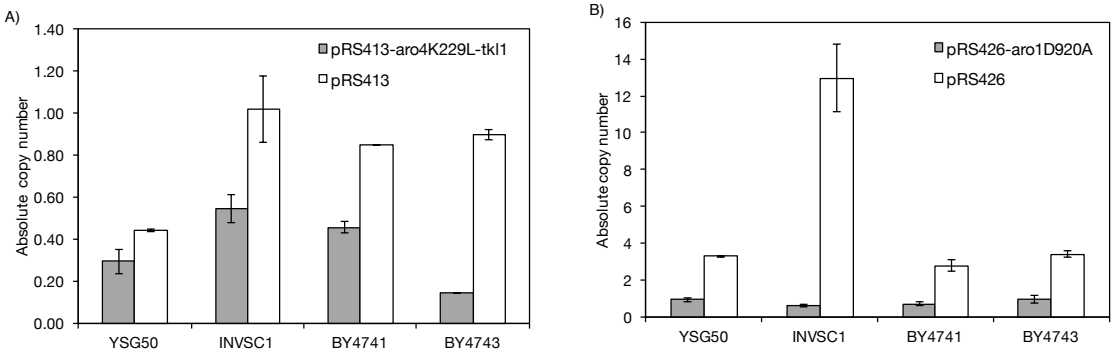


Figure 9: **Absolute copy number of plasmid in strains SA0 and SA2.** A) Comparison between plasmids pRS413 and pRS413-*aro4*$_{K229L}$-*tkl1*. B) Copy number comparison between plasmids pRS426 and pRS426-*aro1*$_{D920A.}$

To test this hypothesis, we constructed a new plasmid encoding the three genes on the pRS413 backbone (pRS413-*aro1*$_{D920A}$-*aro4*$_{K229L}$-*tkl1*). To maintain the same selective nutrients as used previously, this plasmid was transformed along with an empty pRS426 vector yielding the group SA3. As expected, all the strains in group SA3 showed increases in the SA titers. A 2-fold increase was observed for strains YSG50 and BY4741, and the most significant increase occurred in strain BY4743-SA3, where the SA titer was elevated 5.5-fold over the corresponding. The strain INVSc1-SA3 showed very similar titers when compared to INVSc1-SA2 (358 and 350 mgL^{-1}), indicating that such strains might have reached the maximum flux attainable into the aromatic amino acid pathway with the provided manipulations (**Figure 10**). In parallel, we reconstructed a new plasmid encoding all three genes on the pRS426 backbone (pRS426-*aro1*$_{D920A}$-*aro4*$_{K229L}$-*tkl1*), and co-transformed it with the empty pRS413 into all strains, yielding SA producing groups SA4 (**Table 1**). Very interestingly, a dramatic decrease in titers was observed in all strains in this group. In the case of INVSc1, BY4741, and BY4743, the production dropped 94%, 96%, and 83%, respectively, whereas the strain YSG50 showed no SA accumulation (**Figure 10**). This result suggested that overexpressing *aro4*$_{K229L}$ and *tkl1* excessively added a new layer of stress that negatively affected cellular metabolism.

Contrary to our findings, Karim et al (2013) showed that using histidine and uracil as selection markers result in similar growth rates and plasmid copy number when comparing strains BY4741 and BY4743. In our study, we did observe major discrepancies that can be attained to 1) co-expression of two distinct episomal plasmids and 2) constitutive overexpression of structural genes. Although the four wildtype strains studied here have slight differences in their auxotrophic markers, no correlation with their capacity to accumulate shikimate can be observed. For instance, the only major auxotrophic difference between strains

YSG50 and INVSc1 is the deletion of *ade1* in the former, yet these strains have the most evident disparity in accumulation of shikimate. Similarly, supplementation of L-trp in the growth medium brings the strains INVSc1 and BY4743 to the same auxotrophic level, but their capacity to produce shikimic is considerably different.

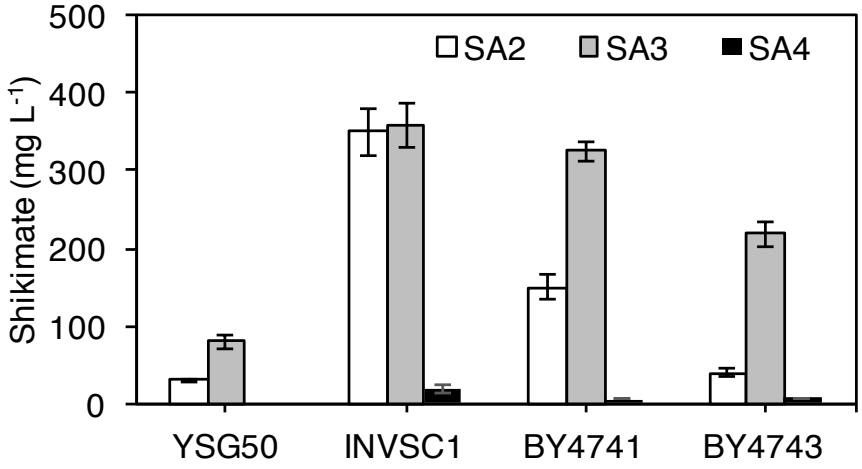


Figure 10: Accumulation of SA in the strains of groups SA2, SA3, and SA4. All strains showed an increased production of SA when the genes were encoded in the plasmid pRS413 (group SA3). Conversely, incorporating all three genes into the backbone of pRS426 (group SA4) resulted in extremely low production of SA.

Balancing gene expression for efficient production of chemicals is a common task carried out in metabolic engineering research (Du et al., 2012; Kim et al., 2013). The results presented here reaffirm that pathway balancing is an obligated step for the construction of efficient microbial factories. Moreover, the pipeline for optimization should also consider the selection of the host strain and tailoring the plasmid selection according to the host. Recently, it has been recognized that yeast strains can auto-adjust plasmid copy number in accordance to the burden of excessive enzyme expression (Carquet et al., 2015). Our study showed that four independent, yet commonly used *S. cerevisiae* strains had disparate responses to simple

metabolic manipulations, which was correlated to their ability to balance the introduced plasmids. A careful inspection of the behavior of each strain brought valuable insights into how to properly engineer the carbon entrance into the aromatic amino acid pathway.

5. Conclusion

Altogether, we were able to observe high-level accumulation of SA in three of the four strains engineered in this work. Having four *S. cerevisiae* strains with different capacity to produce SA, allowed us to implement a strain-dependency approach to discern the appropriate way to engineer our pathway of interest. The higher production of strains in the group SA3 was a result of rationally engineering the entrance of carbon into the aromatic amino acid biosynthetic pathway. This work emphasizes the significance of strain and plasmid selection in metabolic engineering efforts and provides insights to optimize the production of aromatic compounds in yeast.

6. Bibliography

Bochkov, D., Sysolyatin, S., Kalashnikov, A., Surmacheva, I., 2012. Shikimic acid: review of its analytical, isolation, and purification techniques from plant and microbial sources. J Chem Biol. 5**,** 5-17.

Bohlmann, J., Lins, T., Martin, W., Eilert, U., 1996. Anthranilate synthase from *Ruta graveolens*. Duplicated AS alpha genes encode tryptophan-sensitive and tryptophan-insensitive isoenzymes specific to amino acid and alkaloid biosynthesis. Plant Physiol. 111**,** 507-14.

Braus, G. H., 1991. Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway. Microbiol Rev. 55**,** 349-70.

Caesar, R., Palmfeldt, J., Gustafsson, J. S., Pettersson, E., Hashemi, S. H., Blomberg, A., 2007. Comparative proteomics of industrial lager yeast reveals differential expression of the *cerevisiae* and non- *cerevisiae* parts of their genomes. Proteomics. 7**,** 4135-4147.

Carquet, M., Pompon, D., Truan, G., 2015. Transcription interference and ORF nature strongly affect promoter strength in a reconstituted metabolic pathway. Front Bioeng Biotechnol. 3**,** 21.

Chandran, S. S., Yi, J., Draths, K. M., von Daeniken, R., Weber, W., Frost, J. W., 2003. Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. Biotechnol Prog. 19**,** 808-14.

Chen, K., Dou, J., Tang, S., Yang, Y., Wang, H., Fang, H., Zhou, C., 2012. Deletion of the aroK gene is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli*. Bioresour Technol. 119**,** 141-147.

Cui, Y. Y., Ling, C., Zhang, Y. Y., Huang, J., Liu, J. Z., 2014. Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. Microb Cell Fact. 13**,** 21.

Curran, K. A., Leavitt, J. M., Karim, A. S., Alper, H. S., 2012. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng. 15**,** 55-66.

Du, J., Yuan, Y., Si, T., Lian, J., Zhao, H., 2012. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Res. 40**,** e142.

Evans, J., Commercial Amino Acids. Vol. BIO007K. BCC Research Wellesley MA USA, 2014, pp. 214.

Facchini, P. J., 2001. Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. Annu Rev Plant Physiol Plant Mol Biol. 52**,** 29-66.

Feng, X., Zhao, H., 2013a. Investigating glucose and xylose metabolism in *Saccharomyces cerevisiae* and *Scheffersomyces stipitis* via ^{13}C metabolic flux analysis. AIChE J. 59**,** 3195-3202.

Feng, X., Zhao, H., 2013b. Investigating xylose metabolism in recombinant *Saccharomyces cerevisiae* via ^{13}C metabolic flux analysis. Microb Cell Fact. 12**,** 114.

Feng, X., Zhuang, W. Q., Colletti, P., Tang, Y. J., 2012. Metabolic pathway determination and flux analysis in nonmodel microorganisms through ^{13}C-isotope labeling. Methods Mol Biol. 881**,** 309-30.

Fry, B., Zhu, T., Domach, M. M., Koepsel, R. R., Phalakornkule, C., Ataai, M. M., 2000. Characterization of growth and acid formation in a *Bacillus subtilis* pyruvate kinase mutant. Appl Environ Microbiol. 66**,** 4045-9.

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., Smolke, C. D., 2015. Complete biosynthesis of opioids in yeast. Science. 349**,** 1095-100.

Garcia-Albornoz, M., Thankaswamy-Kosalai, S., Nilsson, A., Varemo, L., Nookaew, I., Nielsen, J., 2014. BioMet Toolbox 2.0: genome-wide analysis of metabolism and omics data. Nucleic Acids Res. 42**,** W175-81.

Hansen, E. H., Moller, B. L., Kock, G. R., Bunner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S., Hansen, J., 2009. *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). Appl Environ Microbiol. 75**,** 2765.

Iraqui, I., Vissers, S., Cartiaux, M., Urrestarazu, A., 1998. Characterisation of *Saccharomyces cerevisiae* ARO8 and ARO9 genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily. Mol Genet Genomics. 257**,** 238-48.

Karim, A. S., Curran, K. A., Alper, H. S., 2013. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. FEMS Yeast Res. 13**,** 107-116.

Kawai, S., Hashimoto, W., Murata, K., 2010. Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. Bioeng Bugs. 1**,** 395-403.

Kim, B., Du, J., Eriksen, D. T., Zhao, H., 2013. Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. Appl Environ Microbiol. 79**,** 931-41.

Kim, B., Park, H., Na, D., Lee, S. Y., 2014. Metabolic engineering of *Escherichia coli* for the production of phenol from glucose. Biotechnol J. 9**,** 621-9.

Kim, S. R., Xu, H., Lesmana, A., Kuzmanovic, U., Au, M., Florencia, C., Oh, E. J., Zhang, G., Kim, K. H., Jin, Y. S., 2015. Deletion of PHO13, encoding haloacid dehalogenase type IIA phosphatase, results in upregulation of the pentose phosphate pathway in *Saccharomyces cerevisiae*. Appl Environ Microbiol. 81**,** 1601-9.

Koopman, F., Beekwilder, J., Crimi, B., Van Houwelingen, A., Hall Robert, D., Bosch, D., Van Maris Antonius, J., Pronk Jack, T., Daran, J.-M., 2012. *De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. Microb Cell Fact. 11**,** 155.

Kummel, A., Ewald, J. C., Fendt, S. M., Jol, S. J., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., Heinemann, M., 2010. Differential glucose repression in common yeast strains in response to HXK2 deletion. FEMS Yeast Res. 10**,** 322-32.

Lee, K., Hahn, J. S., 2013. Interplay of Aro80 and GATA activators in regulation of genes for catabolism of aromatic amino acids in *Saccharomyces cerevisiae*. Mol Microbiol. 88**,** 1120-34.

Leonard, E., Lim, K. H., Saw, P. N., Koffas, M. A., 2007. Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. Appl Environ Microbiol. 73**,** 3877-86.

Liu, D. F., Ai, G. M., Zheng, Q. X., Liu, C., Jiang, C. Y., Liu, L. X., Zhang, B., Liu, Y. M., Yang, C., Liu, S. J., 2014. Metabolic flux responses to genetic modification for shikimic acid production by *Bacillus subtilis* strains. Microb Cell Fact. 13**,** 40.

Maeda, H., Dudareva, N., 2012. The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol. 63**,** 73-105.

McQuade, B., Blair, M., 2015. Influenza treatment with oseltamivir outside of labeled recommendations. Am J Health Syst Pharm. 72**,** 112-6.

Minami, H., Kim, J. S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., Sato, F., 2008. Microbial production of plant benzylisoquinoline alkaloids. Proc Natl Acad Sci U S A. 105**,** 7393-8.

Moriya, H., Shimizu-Yoshida, Y., Kitano, H., 2006. *In vivo* robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. PLoS Genet. 2**,** e111.

Nilsson, U., Meshalkina, L., Lindqvist, Y., Schneider, G., 1997. Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis. J Biol Chem. 272**,** 1864-9.

Niu, W., Draths, K. M., Frost, J. W., 2002. Benzene-free synthesis of adipic acid. Biotechnol Prog. 18**,** 201-11.

Patnaik, R., Liao, J. C., 1994. Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. Appl Environ Microbiol. 60**,** 3903-8.

Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., Nielsen, J., 2015. Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. Metab Eng. 31**,** 181-188.

Schaaff-Gerstenschlager, I., Mannhaupt, G., Vetter, I., Zimmermann, F. K., Feldmann, H., 1993. TKL2, a second transketolase gene of *Saccharomyces cerevisiae*. Cloning, sequence and deletion analysis of the gene. Eur J Biochem. 217**,** 487-92.

Shao, Z., Luo, Y., Zhao, H., 2012. DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*. Methods Mol Biol. 898**,** 251-62.

Shao, Z., Zhao, H., 2013. Construction and engineering of large biochemical pathways via DNA assembler. Methods Mol Biol. 1073**,** 85-106.

Shin, S. Y., Jung, S. M., Kim, M. D., Han, N. S., Seo, J. H., 2012. Production of resveratrol from tyrosine in metabolically engineered *Saccharomyces cerevisiae*. Enzyme Microb Technol. 51**,** 211-6.

Sprague, G. F., Jr., 1977. Isolation and characterization of a *Saccharomyces cerevisiae* mutant deficient in pyruvate kinase activity. J Bacteriol. 130**,** 232-41.

Strucko, T., Magdenoska, O., Mortensen, U. H., 2015. Benchmarking two commonly used *Saccharomyces cerevisiae* strains for heterologous vanillin-β-glucoside production. Metab Engin Commun. 2**,** 99-108.

Suastegui, M., Matthiesen, J. E., Carraher, J. M., Hernandez, N., Rodriguez Quiroz, N., Okerlund, A., Cochran, E. W., Shao, Z., Tessonnier, J. P., 2016. Combining Metabolic Engineering and Electrocatalysis: Application to the Production of Polyamides from Sugar. Angew Chem Int Ed Engl. 55**,** 2368-73.

Thodey, K., Galanie, S., Smolke, C. D., 2014. A microbial biomanufacturing platform for natural and semisynthetic opioids. Nat Chem Biol. 10**,** 837-44.

Wahl, S. A., Dauner, M., Wiechert, W., 2004. New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol Bioeng. 85**,** 259-68.

Zhang, H., Pereira, B., Li, Z., Stephanopoulos, G., 2015. Engineering *Escherichia coli* coculture systems for the production of biochemical products. Proc Natl Acad Sci U S A. 112**,** 8266-71.

Zhu, T., Pan, Z., Domagalski, N., Koepsel, R., Ataai, M. M., Domach, M. M., 2005. Engineering of *Bacillus subtilis* for enhanced total synthesis of folic acid. Appl Environ Microbiol. 71**,** 7122-9.

6. Supplementary Information

6.1 Supplementary Tables

Supplementary Table 1: Isotopic labeling patterns of SA2 strains.

		YSG50		INVSc1		BY4741		BY4743	
Amino acid	Labeling pattern	M-57	M-159	M-57	M-159	M-57	M-159	M-57	M-159
Ala	m0	0.51	0.49	0.49	0.47	0.48	0.46	0.51	0.49
	m1	0.32	0.34	0.34	0.35	0.35	0.35	0.32	0.34
	m2	0.03	0.17	0.03	0.18	0.02	0.19	0.03	0.17
	m3	0.14		0.14		0.15		0.14	
Gly*	m0	0.85	0.84	0.79	0.83	0.79	0.82	0.85	0.87
	m1	0.03	0.16	0.06	0.17	0.05	0.18	0.02	0.13
	m2	0.12		0.15		0.16		0.12	
Val	m0	0.32	0.31	0.27	0.28	0.26	0.26	0.32	0.31
	m1	0.29	0.28	0.31	0.30	0.29	0.27	0.29	0.28
	m2	0.19	0.25	0.21	0.25	0.21	0.31	0.19	0.25
	m3	0.12	0.13	0.12	0.13	0.15	0.13	0.12	0.13
	m4	0.05	0.03	0.07	0.03	0.06	0.04	0.05	0.03
	m5	0.03		0.02		0.03		0.03	
Leu	m0	0.94	0.95	0.96	0.94	0.94	0.95	0.93	0.95
	m1	0.06	0.04	0.04	0.05	0.06	0.04	0.06	0.03
	m2	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	m3:6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ile	m0	0.45	0.27	0.43	0.24	0.41	0.24	0.46	0.29
	m1	0.21	0.26	0.18	0.26	0.22	0.27	0.20	0.26
	m2	0.17	0.26	0.19	0.28	0.18	0.26	0.19	0.25
	m3	0.11	0.15	0.11	0.16	0.09	0.17	0.08	0.14

Supplementary Table 1 continued.

	m4	0.04	0.05	0.05	0.05	0.07	0.06	0.05	0.05
	m5	0.02	0.01	0.03	0.01	0.03	0.01	0.02	0.01
	m6	0.00		0.00		0.01		0.00	
Pro	m0	0.31	0.32	0.24	0.27	0.24	0.29	0.31	0.32
	m1	0.24	0.29	0.29	0.31	0.27	0.29	0.24	0.29
	m2	0.25	0.24	0.28	0.27	0.28	0.27	0.25	0.24
	m3	0.15	0.12	0.13	0.12	0.15	0.12	0.15	0.12
	m4	0.04	0.03	0.06	0.03	0.05	0.03	0.04	0.03
	m5	0.01		0.01		0.01		0.01	
Ser	m0	0.49	0.55	0.55	0.54	0.45	0.47	0.52	0.51
	m1	0.32	0.32	0.29	0.34	0.37	0.39	0.31	0.34
	m2	0.06	0.13	0.06	0.12	0.01	0.14	0.06	0.14
	m3	0.13		0.11		0.17		0.12	
Thr	m0	0.42	0.39	0.41	0.41	0.41	0.43	0.48	0.47
	m1	0.32	0.39	0.37	0.37	0.36	0.35	0.31	0.33
	m2	0.12	0.23	0.07	0.19	0.08	0.18	0.06	0.15
	m3	0.11	0.00	0.11	0.03	0.11	0.04	0.11	0.05
	m4	0.03		0.05		0.03		0.04	
Phe	m0	0.81	0.80	0.64	0.61	0.65	0.63	0.80	0.79
	m1	0.09	0.11	0.11	0.14	0.12	0.14	0.11	0.11
	m2	0.04	0.04	0.06	0.10	0.08	0.10	0.03	0.04
	m3	0.02	0.02	0.07	0.06	0.06	0.05	0.02	0.03
	m4	0.02	0.01	0.04	0.03	0.04	0.03	0.01	0.01
	m5	0.01	0.01	0.04	0.03	0.03	0.03	0.01	0.01
	m6	0.01	0.01	0.01	0.02	0.02	0.02	0.01	0.01
	m7	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00
	m8:9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Asp	m0	0.43	0.45	0.40	0.41	0.40	0.43	0.42	0.43
	m1	0.30	0.34	0.33	0.33	0.37	0.36	0.36	0.33
	m2	0.12	0.17	0.12	0.21	0.06	0.16	0.09	0.19
	m3	0.13	0.04	0.12	0.05	0.12	0.06	0.10	0.05
	m4	0.02		0.03		0.04		0.03	

Supplementary Table 1 continued.

Glu	m0	0.27	0.29	0.23	0.23	0.20	0.23	0.31	0.32
	m1	0.29	0.26	0.28	0.33	0.30	0.31	0.22	0.26
	m2	0.20	0.28	0.27	0.29	0.28	0.29	0.23	0.28
	m3	0.18	0.15	0.15	0.12	0.16	0.12	0.20	0.11
	m4	0.06	0.03	0.05	0.03	0.05	0.04	0.05	0.03
	m5	0.00		0.01		0.01		0.00	
Lys	m0	0.24	0.25	0.17	0.17	0.17	0.17	0.24	0.25
	m1	0.18	0.23	0.20	0.21	0.21	0.22	0.18	0.23
	m2	0.27	0.23	0.24	0.30	0.26	0.31	0.27	0.23
	m3	0.19	0.20	0.23	0.21	0.22	0.20	0.19	0.20
	m4	0.09	0.07	0.12	0.08	0.08	0.08	0.09	0.07
	m5	0.04	0.02	0.03	0.02	0.05	0.01	0.04	0.02
	m6	0.00	-	0.01	-	0.01	-	0.00	-
Tyr	m0	0.66	0.57	0.52	0.52	0.48	0.44	0.66	0.57
	m1	0.10	0.19	0.17	0.17	0.16	0.18	0.10	0.19
	m2	0.09	0.11	0.09	0.13	0.11	0.15	0.09	0.11
	m3	0.05	0.04	0.07	0.07	0.10	0.09	0.05	0.04
	m4	0.05	0.04	0.05	0.04	0.06	0.06	0.05	0.04
	m5	0.03	0.02	0.05	0.04	0.04	0.03	0.03	0.02
	m6	0.02	0.01	0.02	0.02	0.03	0.03	0.02	0.01
	m7	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	m8	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
	m9	0.00	-	0.00	-	0.00	-	0.00	-
Met	m0	0.27	0.33	0.26	0.28	0.82	0.88	0.27	0.33
	m1	0.34	0.34	0.30	0.36	0.15	0.09	0.34	0.34
	m2	0.19	0.22	0.23	0.23	0.02	0.00	0.19	0.22
	m3	0.11	0.08	0.12	0.11	0.00	0.02	0.11	0.08
	m4	0.09	0.02	0.07	0.02	0.01	0.00	0.09	0.02
	m5	0.01	-	0.02	-	0.00	-	0.01	-
SA	m0	-	-	0.2193	-	0.2255	-	0.2408	-
	m1	-	-	0.2932	-	0.3219	-	0.3317	-

Supplementary Table 1 continued.

	m2	-	-	0.1675	-	0.124	-	0.1341	-
	m3	-	-	0.153	-	0.1508	-	0.1417	-
	m4	-	-	0.0824	-	0.1074	-	0.078	-
	m5	-	-	0.0846	-	0.0705	-	0.0738	-

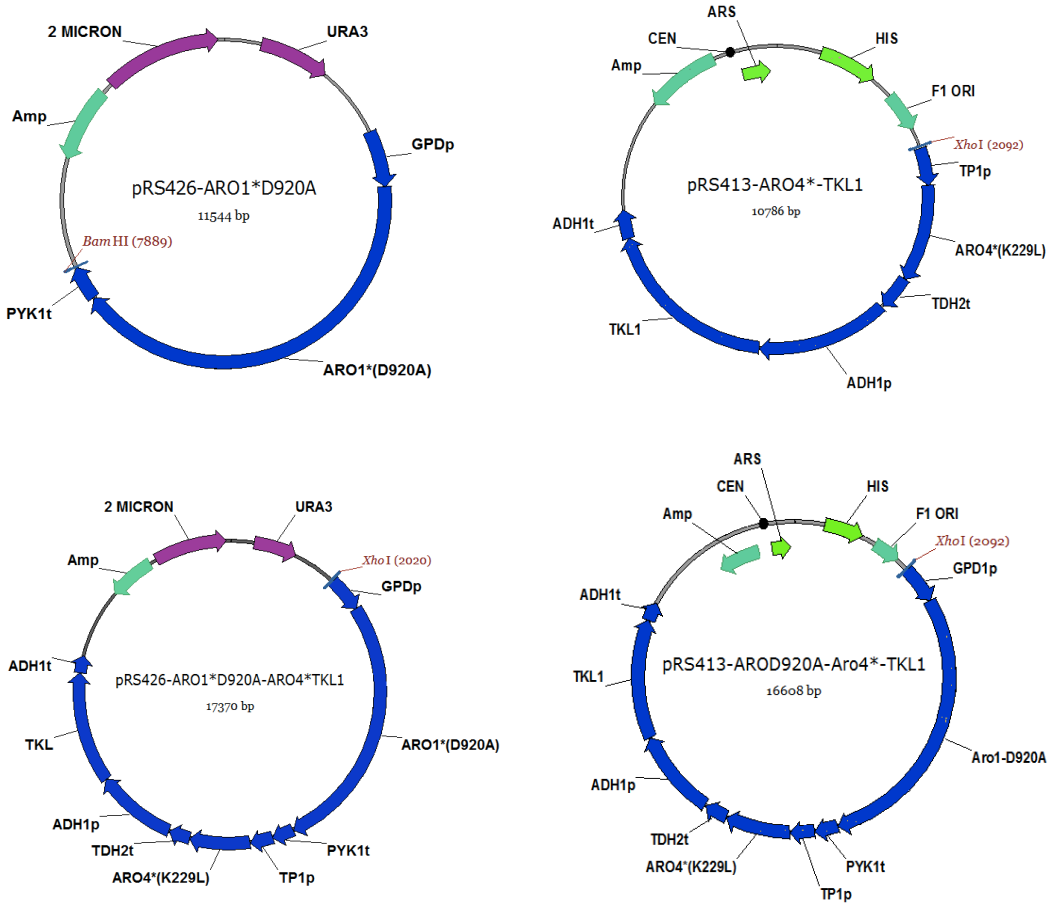
Supplementary Table 2: Reaction model for MFA.

Reaction numeration	Atom transition	Stoichiometry chemistry
Glycolysis pathway		
v1	abcdef <-> abcdef	GLC <-> G6P
v2	abcdef <-> abcdef	G6P <-> F6P
v3	abcdef <-> cba + def	F6P <-> DHAP + G3P
v4	abc <-> abc	DHAP <-> G3P
v5	abc <-> abc	G3P <-> PEP
v6	abc <-> abc	PEP <-> PYRCYT
Pentose phosphate pathway		
v7	abcdef <-> a + bcdef	G6P <-> CO2 + P5P
v8	abcde + fghij <-> fgabcde + hij	P5P + P5P <-> S7P + G3P
v9	abcdefg + hij <-> abchij + defg	S7P + G3P <-> F6P + E4P
v10	abcde + fghi <-> abfghi + cde	P5P + E4P <-> F6P + G3P
Shikimate synthesis		
v11	abc + defg <-> abcdefg	PEP + E4P <-> SA
v12	abc + defghij <-> abcefghij + d	PEP + SA <-> PHE + CO2
Ethanol, acetate, and glycerol formation		
v13	abc <-> bc + a	PYRCYT <-> ACA + CO2
v14	ab <-> ab	ACA <-> ETH
v15	ab <-> ab	ACA <-> ACE
v16	abc <-> abc	G3P <-> GLYC
Formation of AcCoA in the cytosol		
v17	ab <-> ab	ACE <-> ACCOACYT
Anaplerotic reaction (cytosolic)		
v18	abc + d <-> abcd	PYRCYT + CO2 <-> OAACYT
TCA-cycle		
v19	abc <-> bc + a	PYRMIT <-> ACCOAMIT + CO2
v20	abcd + ef <-> dcbafe	OAAMIT + ACCOAMIT <-> ICIT
v21	abcdef <-> abcef + d	ICIT <-> AKG + CO2
v22	abcde <-> bcde + a	AKG <-> FUM + CO2
v23	abcd + efgh <-> abcd + hgfe	FUM + FUM <-> OAAMIT + OAAMIT
Transports		
v24	abcd <-> abcd	OAAMIT <-> OAACY
v25	ab <-> ab	ACCOACYT <-> ACCOAMIT
v26	abc <-> abc	PYRCYT <-> PYRMIT

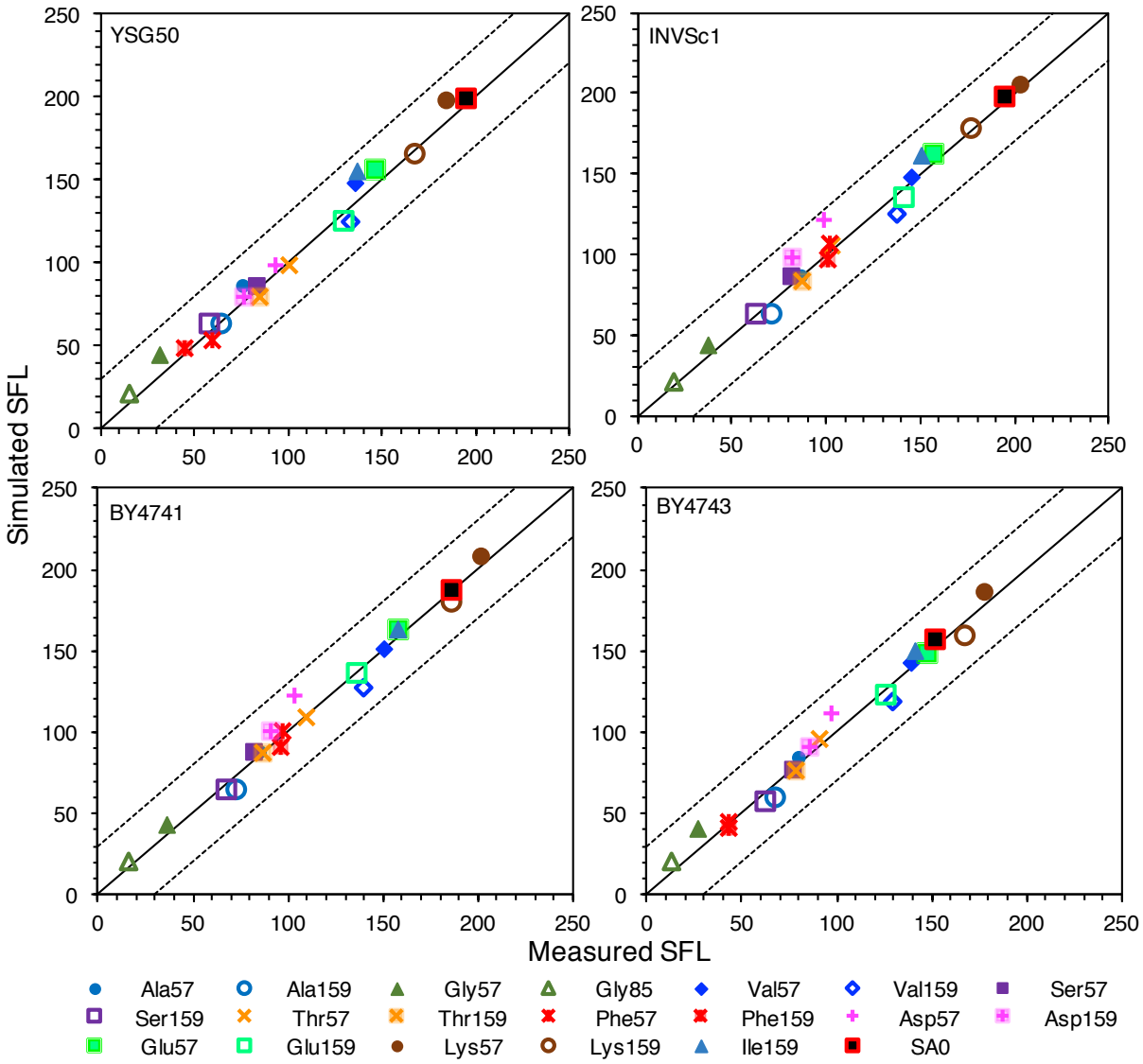
Supplementary Table 2 continued.

Threonine / Serine / Glycine metabolism		
v27	abc <-> abc	G3P <-> SER
v28	abc <-> ab + c	SER <-> GLY + C1
v29	abcd <-> abcd	OAACYT <-> THR
v30	abcd <-> ab + cd	THR <-> GLY + ACA

6.2 Supplementary Figures



Supplementary Figure 1: Plasmids constructed in this work.



Supplementary Figure. 2: Fitting between measured and simulated summed fractional labeling (SFL) of strains in group SA2 (Table 1).

CHAPTER 5

MULTILEVEL ENGINEERING OF THE UPSTREAM AROMATIC AMINO ACID BIOSYNTHESIS MODULE IN *S. CEREVISIAE* FOR HIGH PRODUCTION OF POLYMER AND DRUG AROMATIC PRECURSORS

Authors: Miguel Suástegui[1,3], Chiam Yu Ng[2,3], Anupam Chowdhury[2,3], Wan Sun[3,4], Emma House[3], Costas D. Maranas[2,3], Zengyi Shao[*,1,3,4,5]

[1]Department of Chemical and Biological Engineering, Iowa State University, 618 Bissell Road, Ames, Iowa 50011, United States

[2]Department of Chemical Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, United States

Abstract

A multilevel approach was implemented in *Saccharomyces cerevisiae* to optimize the precursor module of the aromatic amino acid biosynthesis pathway, which is a rich resource for synthesizing a great variety of chemicals ranging from polymer precursor, to nutraceuticals and pain-relief drugs. To facilitate the discovery of novel targets to enhance the pathway flux, we incorporated the computational tool YEASTRACT for predicting novel transcriptional repressors and OptForce strain-design for identifying non-intuitive pathway interventions. The multilevel approach consisted of (i) relieving the pathway from strong transcriptional repression, (ii) removing competing pathways to ensure high carbon capture, and (iii) rewiring precursor pathways to increase the carbon funneling to the desired target. The combination of these interventions led to the establishment of a *S. cerevisiae* strain with shikimic acid (SA) titer reaching as high as 2.48 g L^{-1}, 7-fold higher than the base strain. Further expansion of the platform led to the production of muconic acid (MA) and its intermediate protocatechuic acid together close to 1.6 g L^{-1}. Both SA and MA production platforms demonstrated increases in titer and yield around 300% from the previously reported, highest-producing *S. cerevisiae*

strains. Further examination elucidated the diverged impacts of disrupting the oxidative branch (*ZWF1*) of the pentose phosphate pathway on the titers of desired products belonging to different portions of the pathway. The investigation of other non-intuitive interventions like the deletion of the Pho13p enzyme also revealed the important role of the transaldolase in determining the fate of the carbon flux in the pathways of study. This integrative approach identified novel determinants at both transcriptional and metabolic levels that constrain the flux entering the aromatic amino acid pathway and provided a readily usable platform for engineering the downstream modules in the future towards production of important plant-sourced aromatic secondary metabolites.

1. Introduction

Engineering microbial factories for the production of valuable chemicals often requires the optimization of long metabolic pathways. Splitting the pathways into modules can allow for faster optimization, leading to higher overall yields and titers. Grouping genetic elements permits a more precise analysis of bottlenecks, rate-limiting steps, and metabolic imbalances (Biggs et al., 2014). Moreover, studying metabolic pathways as independent modules facilitates the analysis of transcriptional regulators that may act on the genes within distinct modules, hence enabling a multilevel approach encompassing engineering at both transcriptional and metabolic levels.

Modular approaches for engineering microbes have been implemented in a variety of cases. In *Escherichia coli*, for example, implementing a modular design strategy led to significant increases in the production of total fatty acids (Xu et al., 2013). In *Saccharomyces cerevisiae*, the most recent example was observed for the *de novo* production of benzylisoquinoline alkaloids (BIAs), which required the coordinated overexpression of more than 20 endogenous

and heterologous genes (Galanie et al., 2015). Partitioning the pathway into several modules allowed efficient optimization of precursor and cofactor availability, and reduction of pathway bottlenecks. In fact, the aromatic amino acid pathway from which BIAs are originated is a source of a great variety of chemicals that range from polymer precursors to nutraceuticals and pain-relief drugs (Suástegui and Shao, 2016). Previous metabolic engineering efforts in *S. cerevisiae* have enabled the production of chemicals derived from this pathway including, but not limited to, shikimic acid (SA) (Suástegui et al., 2016a), muconic acid (MA) (Curran et al., 2012; Suástegui et al., 2016b), vanillin (Brochado et al., 2010), L-tyrosine (L-tyr) (Gold et al., 2015), coumaric acid (Rodriguez et al., 2015), and secondary metabolites from the flavonoid and stilbenoid families, as well as BIAs (Galanie et al., 2015; Jiang et al., 2005; Koopman et al., 2012; Li et al., 2015; Trantas et al., 2009; Yan et al., 2005).

One major metabolic engineering strategy employed to unlock the production of derivatives spun off the aromatic amino acid biosynthetic pathway is the overexpression of the feedback insensitive enzymes 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase, Aro3/4p, and chorismate mutase, Aro7p, catalyzing the first committed step in the pathway, and the branching point towards the production of L-tyr and phenylalanine (L-phe), respectively (Brown and Dawes, 1990; Luttik et al., 2008). As shown in **Figure 1**, other strategies include rewiring the pentose phosphate pathway (PPP) for increasing the pool of erythrose-4-phosphate (E4P), and the overexpression of mutant versions of the pentafunctional protein Aro1 to halt the activity at specific subunits for accumulating the desired intermediates (Suástegui and Shao, 2016). Despite the implementation of these strategies, the carbon yields in *S. cerevisiae* still remain low, suggesting the existence of unknown determinants constraining the carbon flux entering the pathway. Therefore, a more comprehensive

investigation needs to be prosecuted to determine novel transcriptional and metabolic targets, which will benefit the design of microbial cell factories for producing many high-value compounds from such a resourceful pathway.

Herein, we have focused the attention on engineering the precursor module of the aromatic amino acid pathway using SA as a reporter product. The target module is composed of the genes in the glycolytic pathway (up to the production of phosphoenolpyruvate, PEP), the PPP (leading to the production of erythrose-4-phosphate, E4P), and the genes in the SA pathway including *ARO2*, *ARO3*, *ARO4*, and the pentafunctional *ARO1* (**Figure 1**). Besides implementing a comprehensive engineering strategy at the metabolic level (*i.e.* deletion or overexpression of structural genes), we identified that the protein Ric1, involved in regulating the efficient localization of *trans*-Golgi network proteins (Bensen et al., 2001), can act as a transcriptional repressor of multiple genes in the aromatic amino acid pathway. Furthermore, the incorporation of *in silico* modeling and pathway analysis led to the discovery of a novel genetic target. Overexpression of ribose-5-phosphate ketol-isomerase, *RKI1*, in addition to other tested interventions, facilitated PPP flux redirection into the aromatic amino acid pathway through E4P. The combination of these novel, multilevel interventions led to the high production of SA, MA, and its intermediate. Such a multilevel intervention strategy thoroughly removes the constraints in the upstream precursor module of the aromatic amino acid pathway, maximize the carbon flux flowing to the downstream, and therefore pave the way for synthesizing high-value molecules from the downstream branches.

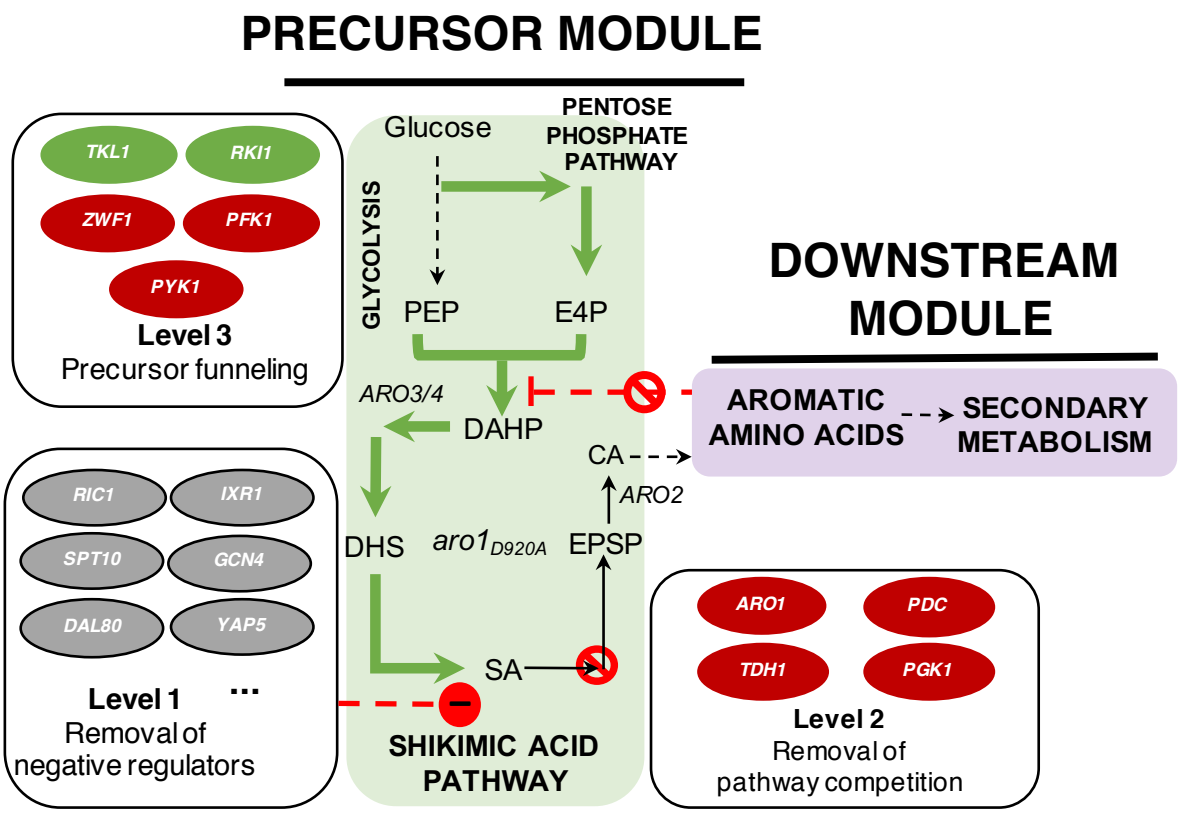


Figure 1: **Modular visualization of the aromatic amino acid biosynthetic pathway.** The multilevel strategy consisted of removing pathway repressors (level 1), removing pathway competition (level 2), and increasing precursor funneling (level 3). Color codes: green, gene overexpressions; red, gene knockouts. Metabolite abbreviations: PEP, phosphoenolpyruvate; E4P, erythrose-4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate-7-phosphate; DHS, 3-dehydroshikimic acid; SA, shikimic acid; EPSP, 5-enolpyruvyl-3-shikimate phosphate; CA, chorismic acid. Gene abbreviations: *TKL1*, transketolase; *RKI1*, ribose-5-phosphate ketol-isomerase; *ZWF1*, glucose-6-phosphate dehydrogenase; *PFK1*, phosphofructokinase; *ARO1*, pentafunctional aromatic enzyme; $aro1_{D920A}$, mutant version of *ARO1* with disrupted shikimate kinase activity; *ARO2*, chorismate synthase, *PDC*, pyruvate decarboxylase; *PGK1*, 3-phophoglycerate kinase; *TDHI1*, glyceraldehyde-3-phosphate dehydrogenase.

2. Materials and Methods

2.1. Strains and media

The strain *S. cerevisiae* BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was used as the microbial platform for the production of SA and MA, as well as for the assembly of plasmids using the DNA Assembler technique (Shao et al., 2012; Shao and Zhao, 2013; Shao and Zhao, 2014). The yeast cultures were maintained in YPD medium (1% yeast extract, 2% peptone, and 2% dextrose). Antibiotic G418 was supplemented at a concentration of 200 μg Ml^{-1} for the growth of the deletion strains carrying the *KanMX* marker from the *MATa* Yeast Knockout Collection (GE Dharmacon, USA). Synthetic complete (SC) media lacking single or multiple nutrients (0.5% ammonium sulfate, 0.16% yeast nitrogen without amino acid and ammonium sulfate, complete amino acid supplement mix, and 2% dextrose) were used to culture the cells after plasmid transformation. The addition of the three aromatic amino acids L-tyr, L-trp (tryptophan), and L-phe was required at a level of 50 mg L^{-1} each, for strains carrying the *ARO1* deletion. For SA or MA production, minimal medium supplemented with 2% dextrose was used, as previously reported (Feng and Zhao, 2013a; Feng and Zhao, 2013b; Suástegui et al., 2016a). The *E. coli* strain BW25141 was used for plasmid propagation. The bacterial transformation was done by the TSS method (Chung et al., 1989) and cultured in Luria-Bertani (LB) broth supplemented with 100 μg Ml^{-1} ampicillin.

2.2. Plasmid construction

The plasmids constructed in this work were assembled *via* DNA assembler (> 4 fragments) (Shao et al., 2012; Shao and Zhao, 2013; Shao and Zhao, 2014) or Gibson assembly (2-3 fragments) (Gibson et al., 2009). The Pcrct plasmids containing the CRISPR/Cas9 system for

gene editing were constructed *via* Golden Gate assembly (Engler et al., 2009; Engler et al., 2008) following the previously reported design (Bao et al., 2014). A list of the plasmids used in this work is displayed in **Table 1**. Transformation of plasmids into yeast was performed following the "quick and dirty" protocol (Amberg et al., 2006). All yeast genetic fragments (promoters, genes, and terminators) were amplified by PCR using the genomic DNA of strain BY4741. For gene knockouts using plasmid Pcrct, gBlocks® containing the donor DNA and the guide RNA were synthesized by IDT DNA (Coralville, IA).

Table 1: List of plasmids used in this study.

Production plasmids	Genetic elements	Reference
pRS413		
pRS413-lowAA	pRS413-ARO1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL1*	This study
pRS413-midAA	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL1*	(Suastegui et al., 2016a)
pRS413-highAA	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL1* / PGK1-*RKI1*	This study
pRS413-midAA_1	pRS413-GPD1-*aro1*$_{D1409A,D920A}$ / TP1-*aro4*$_{K229L}$ /ADH1-*TKL1*	
pRS413-highAA_1	pRS413-GPD1-*aro1*$_{D1409A,D920A}$ / TP1-*aro4*$_{K229L}$ /ADH1-*TKL1* / PGK1-*RKI1*	
pRS425-MA	pRS425- PYK1-*AROZ* / GPD1 – *HQD2* / TEF1 – *AROY*	(Suastegui et al., 2016b)
CRISPR plasmids		
pCRCT		(Bao et al., 2014)
pCRCT_*aro1*		This study
pCRCT_*pho13*		This study
pCRCT_*zwf1*		This study

2.3. Gene deletion using CRISPR/Cas9

The engineered strains with single or multiple gene deletions are listed in **Table 2**. Deletion of the genes *ARO1, PHO13,* and *ZWF1* was performed using the plasmid Pcrct containing the CRISPR/Cas9 system (Bao et al., 2014). The CRISPR RNA (crRNA) cassettes for gene deletion were composed of 100 bp of the donor fragment lacking 8-bp to induce a frame shift and a 20-bp guide sequence (**Supplementary Table 1**). The guide sequences were designed based on the CHOPCHOP web tool (Montague et al., 2014). After plasmid transformation (1 μg), the yeast cells were grown for two days in liquid SC-Uracil medium. After this period, 200 μL of culture was transferred to fresh liquid medium and incubated for two more days. Finally, 100 μL of 10^4–fold diluted cells were plated on selective medium. Individual colonies were picked and inoculated in YPD plus 5-fluoroorotic acid and grown for 2 days to ensure plasmid loss. Genomic DNA was extracted by two cycles of freeze/boil and the supernatant was used as a template for diagnostic PCR. The PCR product was recovered from agarose gel and sent for Sanger sequencing at the DNA facility at Iowa State University.

Table 2. Knockout *S. cerevisiae* strains.

Strain	Description	Genotype	Source
	BY4741		
AA_001	BY4741 *ric1*Δ	*ric1*Δ*::KanMX*	GE Dharmacon
AA_002	BY4741 *aro1*Δ	*aro1*Δ*::KanMX*	GE Dharmacon
AA_003	BY4741 *ric1*Δ *aro1*Δ	*ric1*Δ*::KanMX* , *aro1*Δ	This study
AA_004	BY4741 *zwf1*Δ	*zwf1*Δ*::KanMX*	GE Dharmacon
AA_005	BY4741 *pho13*Δ *aro1*Δ *ric1*Δ	*ric1*Δ*::KanMX, aro1*Δ, *pho13*Δ	This study
AA_006	BY4741 *zwf1*Δ *aro1*Δ *ric1*Δ	*ric1*Δ*::KanMX, aro1*Δ, *zwf1*Δ	This study

2.4. Analysis of transcriptional repressors

The genes *ARO1*, *ARO2*, *ARO3*, and *ARO4* were chosen for analysis of TFs using the YEASTRACT website (Teixeira et al., 2014). The search for TFs by target gene was performed in the Regulatory Associations tool by filtering the regulation type to only those acting as repressors (referred as 'inhibitors' on the website) with 'evidence of DNA binding plus expression'. This resulted in a list of 66 TFs, from which only 21 were selected for further analysis based on evidence of knockout viability and availability in the *MATa* yeast library (**Table 3**).

Table 3: List of the transcription factors studied in this work. Description collected from the yeast genome online database (http://www.yeastgenome.org/)

Gene	Systematic name	Description
STP1	YDR463W	Transcriptional activator of amino acid permease genes
YAP1	YML007W	Basic leucine zipper transcription factor; required for oxidative stress tolerance
AGP1	YCL025C	Low-affinity amino acid permease with broad substrate range
ZAP1	YJL056C	Zinc-regulated transcription factor
CBF1	YJR060W	Centromere Binding Factor
TGL3	YMR313C	Triacylglycerol Lipase
GAL4	YPL248C	DNA-binding transcription factor required for activating GAL genes
RPN4	YDL020C	Transcription factor that stimulates expression of proteasome genes
NRG1	YDR043C	Transcriptional repressor; recruits the Cyc8p-Tup1p complex to promoters
OAF1	YAL051W	Oleate-Activated transcription factor
AAD3	YCR107W	Putative aryl-alcohol dehydrogenase
MIG3	YER028C	Transcriptional regulator with a major role in catabolite repression and ethanol
YAP5	YIR018W	Basic leucine zipper iron-sensing transcription factor; involved in diauxic shift
CST6	YIL036W	Basic leucine zipper transcription factor involved in stress-responsive regulatory network
ESC2	YDR363W	Establishment of silent chromatin
YRM1	YOR172W	Zinc finger transcription factor involved in multidrug resistance
ARR1	YPR199C	Transcriptional activator required for transcription of genes involved in resistance to arsenic compounds.
IXR1	YKL032C	Transcriptional repressor that regulates hypoxic genes during normoxia
DAL80	YKR034W	Negative regulator of genes in multiple nitrogen degradation pathways
SPT10	YJL127C	Histone H3 acetylase with a role in transcriptional regulation
RIC1	YLR039C	Protein involved in retrograde transport to the cis-Golgi network

Table 3 continued

PHO2	YDL106C	Homeobox transcription factor; regulatory targets include genes involved in phosphate metabolism
PHO3	YBR092C	Constitutively expressed acid phosphatase similar to Pho5p
PHO4	YFR034C	Basic helix-loop-helix transcription factor, activates transcription cooperatively with Pho2p in response to phosphate limitation
PHO5	YBR093C	Repressible acid phosphatases that also mediates extracellular nucleotide-derived phosphate hydrolysis
PHO8	YDR481C	Repressible vacuolar alkaline phosphatase; regulated by levels of Pi
PHO13	YDL236W	Conserved phosphatase acting as a metabolite repair enzyme
PHO80	YOL001W	Regulates the response to phosphate limitation and stress-dependent calcium signaling
PHO84	YML123C	High-affinity Pi transporter; also low-affinity manganese transporter
PHO88	YBR106W	Probable membrane protein; involved in phosphate transport
PHO90	YJL198W	Low-affinity phosphate transporter
PHO92	YDR374C	Posttranscriptional regulator of phosphate metabolism

2.5. Computational modeling

The optimization-based OptForce algorithm (Ranganathan et al., 2010) was implemented to identify target reaction-level interventions leading to the overproduction of SA and MA individually. In particular, the objective of the computational study was to identify key motifs in the intervention strategies that were conserved, and those that differed due to cofactor or energy equivalent requirements, when attempting to overproduce two metabolites in the same branch of metabolism (*i.e.* aromatic amino acid pathway). The genome-scale metabolic model *i*AZ900 (Zomorrodi and Maranas, 2010) was used to simulate the metabolic flux profiles in *S. cerevisiae* for both target molecules. For the case of SA production, the base model was modified by adding the SA exchange and transport reactions. Likewise, for simulation of MA production, the MA production pathway was included in the model by adding DHS

dehydratase (SKHL), PCA decarboxylase (PCC) and catechol 1,2-dioxygenase (CATO) reactions, along with the corresponding exchange and transport reactions (**Supplementary Table 2**). All simulations were performed in the aerobic minimal media with glucose as the sole carbon substrate mimicking the experimental fermentation conditions. Details of simulation conditions and regulation are described in the Supplementary Information. The CPLEX optimization software was used to solve the mixed integer optimization programming problems to optimality, and was accessed through the General Algebraic Modeling System optimization package. Note that the reaction-level intervention strategy suggested by OptForce was converted to gene-level suggestion using the gene-protein-reaction relationship information mined from *i*AZ900 model, as well as from the most recent curation of *S. cerevisiae* metabolism reported in literature (Chowdhury et al., 2015).

2.6. Fermentation and strain characterization

Single colonies were picked from transformation plates and inoculated into 3 mL of SC-Histidine or SC-Histidine-Leucine for seed culturing of the SA and MA production strains, respectively. After overnight growth, 30 μL of the saturated cultured was transferred into 3 mL of minimal medium lacking the corresponding amino acids. 1 mL of samples were collected after 72 h and centrifuged at 5,000 rpm for 5 min; the supernatant was stored at -20 °C for chromatographic analysis.

SA and DHS were analyzed by High Performance Liquid Chromatography (HPLC) (Waters, Milford, MA) with an Aminex HPX-87H column (300×7.8 mm) (Bio-Rad, Hercules, CA) as previously reported (Gao et al., 2016). An ACQUITY Ultra Performance Liquid Chromatography (UPLC) with a BEH-C18 column (Waters) as the stationary phase was used

to analyze MA and PCA. The mobile phase program was implemented as described previously (Suástegui et al., 2016b). Standard curves of PCA (MP Biomedicals, Santa Clara, CA), SA, DHS, and MA (Sigma-Aldrich, St. Louis) with authentic compounds were generated for metabolite quantification.

3. Results

3.1. Identification of a novel transcriptional regulator

SA was selected as the reporter molecule to track the entrance of carbon into the precursor module. Initially, to study the transcriptional regulation upon this module, we focused the attention to the *ARO1* promoter as it controls the expression of the pentafunctional gene *ARO1*, the main core of the module. **Table 1** shows the plasmid design based on the previously reported platform in *S. cerevisiae* for production of SA (Suástegui et al., 2016a). The promoter controlling the expression of the mutant gene $aro1_{D920A}$ was switched from the strong constitutive *GPD1* promoter (glycerol-3-phosphate dehydrogenase, strain SA2) to the native *ARO1* promoter, yielding plasmid pRS413-lowAA (strain SA1, **Table 4 and Figure 2a**). The production after 3 days of fermentation reached 176.27 ± 0.50 mg L^{-1} of SA (**Figure 2b**), a reduction of almost 46% compared to strain SA2. This indicated that, indeed, the two promoters had different strengths, presumably due to distinct responses to transcriptional regulatory elements. Hence we proceeded to investigate which specific transcription factors (TFs) could be involved in the regulation of the upstream precursor module of the aromatic amino acid biosynthesis, including *ARO1*, *ARO2*, *ARO3*, and *ARO4* (**Figure 1**), as a direct approach to discerning the cause of the repressive regulation of carbon flux into the pathway.

A list of TFs with reported evidence of acting as repressors under different environmental conditions was retrieved from the Yeast Search for Transcriptional Regulators and Consensus Tracking (YEASTRACT) website (Teixeira et al., 2014). YEASTRACT is a repository of over 200,000 regulatory associations between transcription factors and target genes in *S. cerevisiae*, developed based on more than 1300 references. The list contained 66 unique TFs from which 21 were selected for further analysis according to the criteria described in the Materials and Methods section (**Table 3**).

Table 4: Shikimic acid producing strains.

Strain	Genotype	Plasmid	Parent
SA1	BY4741	pRS413-ARO1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	BY4741
SA2	BY4741	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	BY4741
SA3	*ric1*Δ*::KanMX*	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	AA_001
SA4	*aro1*Δ*::KanMX*	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	AA_002
SA5	*ric1*Δ*::KanMX , aro1*Δ	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	AA_003
SA6	*zwf1*Δ*::KanMX*	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	AA_004
SA7	*ric1*Δ*::KanMX, aro1*Δ*, pho13*Δ	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL*	AA_005
SA8	*ric1*Δ*::KanMX, aro1*Δ	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL* / PGK1-*TAL*	AA_003
SA9	*ric1*Δ*::KanMX , aro1*Δ	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL* PGK1-*RKI1*	AA_003
SA10	*ric1*Δ*::KanMX, aro1*Δ*, pho13*Δ	pRS413-GPD1-*aro1*$_{D920A}$ / TP1-*aro4*$_{K229L}$ / ADH1-*TKL* PGK1-*RKI1*	AA_005

3.2. Removal of competing pathways

The precursor module in the aromatic amino acid pathway can be exploited for the production of molecules such as drug precursors (*e.g.* SA), polymer precursors (*e.g.* MA), and flavoring agents (*e.g.* vanillin). It is evident that to divert more flux towards the target molecules, deletion of the competing pathways (*i.e.* the biosynthesis of L-tyr, L-trp, and L-phe) is required. In fact, OptForce also suggested the down-regulation of shikimate kinase (SHKK) or 3-phosphoshikimate 1-carboxyvinyltransferase (PSCVT) to restrict flux drainage towards biomass precursors and competitive products during SA overproduction (**Figure 3**). This intervention can be achieved by knocking out the gene *ARO1*, which catalyzes the five intermediate steps in the SA pathway, in combination with overexpression of *ARO1* variant that maintains the catalytic functions of the first three conversions from DAHP to SA. The strain BY4741 Δ*aro1* overexpressing $aro1_{D920A}$, *TKL1*, and $aro4_{K229L}$ (Strain SA4), reached a titer of 606.0 ± 4.8 mg L^{-1} SA (**Figure 4**). By combining the deletions of *RIC1* and *ARO1* (Strain SA5), the production increased to 800.3 ± 26.8 mg L^{-1} SA, which represented a 4.5-fold increase compared to the initial strain SA1 (**Figure 4**).

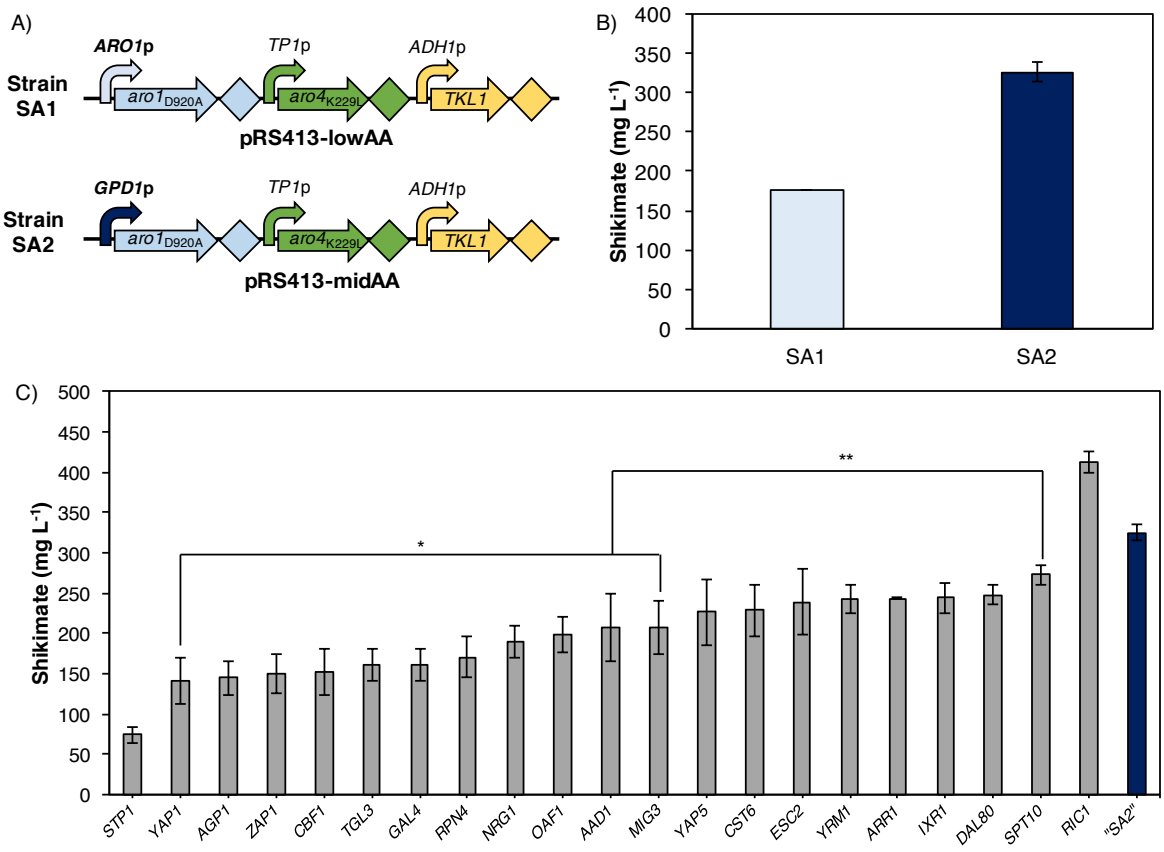


Figure 2: Engineering the production of SA at the transcriptional level. A) Minimal genetic cassettes for production of SA. The transcription of the mutant *aro1*$_{D920A}$ was controlled by the native ARO1p or the constitutive GPD1p. B) Production of SA from strains with different promoters for expression of *aro1*$_{D920A}$. C) Fermentation results from TF KO strains. "SA2" represents the control strain. The variation is represented by the standard deviation from three biological replicates. Statistical different groups are depicted by * and **.

3.3. Upstream manipulations for increased precursor availability

3.3.1. Clarification of the role of oxidative PPP

The next level of engineering was directed towards enabling higher availability of the precursor E4P. Previously, we have identified that between the two precursors PEP and E4P, the latter is the rate limiting one constraining the entrance flux into aromatic amino acid biosynthesis pathway in *S. cerevisiae* (Suástegui et al., 2016a). We began with the deletion of the gene *ZWF1* encoding glucose-6-phosphate dehydrogenase (G6PDH2), a strategy that has been previously implemented for producing MA and L-tyr. The combination of this deletion with the overexpression of transketolase (*TKL1*), has been demonstrated to force the carbon flux to enter through the non-oxidative branch of the pentose phosphate pathway thereby enhancing the pool of E4P (Curran et al., 2012; Gold et al., 2015).

Deletion of *ZWF1* (strain SA6) caused a reduction in SA production to 213.8 ± 6.75 mg L^{-1}, but the total titer of 3-dehydroshikimic acid (DHS) and SA (676.81 mg L^{-1}) increased ~2.1-fold when compared to strain SA2 (**Figure 4**). The reaction catalyzed by G6PDH2 is the main source for generating reducing power in the form of NADPH, which in turn serves as the cofactor for the SA dehydrogenase subunit of Aro1 protein. A reduction in the NADPH pool directly translated into a lower catalytic conversion of the SA dehydrogenase, resulting in a higher accumulation of DHS (Zhang et al., 2015). Very interestingly, this result was consistent with our computational study. In contrary to the prediction of Genetic Design by Local Search (GDLS) (Gold et al., 2015), OptForce did not suggest the knockout of *ZWF1* because SA production requires the cofactor NADPH. Knocking out *ZWF1 in silico* reduced the theoretical

maximum yield of SA by 4.8% (from 64.0 mmol g DW^{-1} h^{-1} to 60.9 mmol g DW^{-1} h^{-1}) because metabolic flux has to be drained towards competitive metabolic pathways such acetaldehyde dehydrogenase to supply the required NADPH.

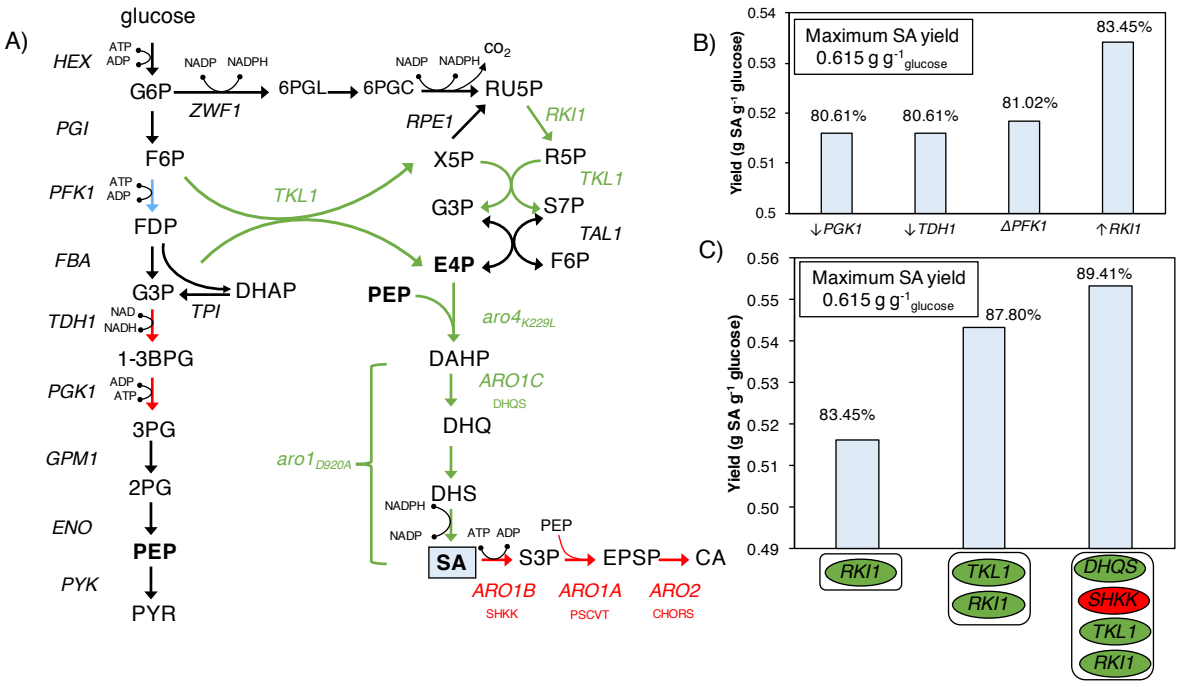


Figure. 3: Metabolic interventions predicted using OptForce. A) Simplified map of the central carbon metabolism depicting the upstream pathway (glycolysis and PPP) leading towards the aromatic amino acid pathway. B) The minimum guaranteed yield achievable by downregulation (↓), deletion (Δ), or overexpression (↑) of the selected novel genes. C) The intervention strategies predicted by OptForce for the overproduction of SA. The overexpression of the genes *RKI1*, *TKL1*, $aro1_{D920A}$ (DHQS), in combination with the deletion of *ARO1* (SHKK), led to a minimum guaranteed SA yield equivalent to 89.41% of the maximum theoretical yield. Green, red, and blue arrows (and circles) represent overexpression, downregulation and knockout of genes, respectively

Deletion of *ZWF1* (strain SA6) appeared to have a similar effect to that of deleting *ARO1* (strain SA4), with accumulation of DHS and a slight reduction (11%) of the overall titer (MA plus DHS). Despite the fact that the deletion of *ZWF1* increases the supply of precursor E4P with a concomitant increase in flux into the SA pathway (observed by comparing strains SA2 vs. SA6), the resulting loss of NADPH cofactor regeneration limits the production of SA and the subsequent metabolites downstream of SA. On a side note, replenishment of cytosolic NADPH by other reactions (*e.g.* aldehyde dehydrogenase *ALD6*) could potentially reactivate the conversion of DHS to SA, but could cause a loss of carbon towards the formation of undesirable byproduct acetate. Based on this study, we clarified that the deletion of *ZWF1* should not be included as a strategy to improve the carbon flux into aromatic amino acid pathway if the target products are downstream of DHS unless NADPH can be efficiently replenished

3.3.2. Clarification of the role of transaldolase

Previous studies have shown that deletion of the haloacid dehalogenase encoded by *PHO13* can increase the transcription rates of the genes in the PPP (Kim et al., 2015). We hypothesized that such manipulation could increase the pool of E4P that enters the aromatic amino acid biosynthesis pathway. Accordingly, we deleted *PHO13* as a new alternative to rewiring the PPP to enhance the production of pathway precursor. However, the strain SA7 (Δ*pho13*) accumulated 731 ± 44 mg L^{-1} of SA, which represented an 11% reduction compared to strain SA5 (Δ*ric1* Δ*aro1*) (**Figure 4**).

Re-inspecting the mechanism of gene up-regulation in a Δ*pho13* strain (Kim et al., 2015) led to the hypothesis that the decrease in SA production could be due to the reverse of carbon flux into glycolysis rather than into the aromatic amino acid pathway. In a Δ*pho13* background,

transaldolase (*TAL1*) has a 3.5-fold increase in expression (Kim et al., 2015). However, the kinetic parameters (K_m) of this transaldolase suggest that the favored direction of the reaction is towards consuming E4P (Tsolas and Horecker, 1973), which is the opposite of what is desired. In fact, the xylose utilization rate in *S. cerevisiae* strains can be enhanced by overexpressing *TAL1* (Vilela Lde et al., 2015), indicating that the transaldolase is pivotal in connecting the flux from PPP to glycolysis. To prove this, we overexpressed *TAL1* in strain SA5 (resulting in strain SA8), leading to a SA titer of 709.4 ± 45 mg L^{-1} (**Figure 4**); no statistical difference was observed when compared to strain SA7 (Student's *t*-test, $p > 0.05$). These results supported our hypothesis that the drop in SA production in strain SA7 was a result of carbon being drained to glycolysis presumably due to the increased expression of *TAL1* when *PHO13* was deleted.

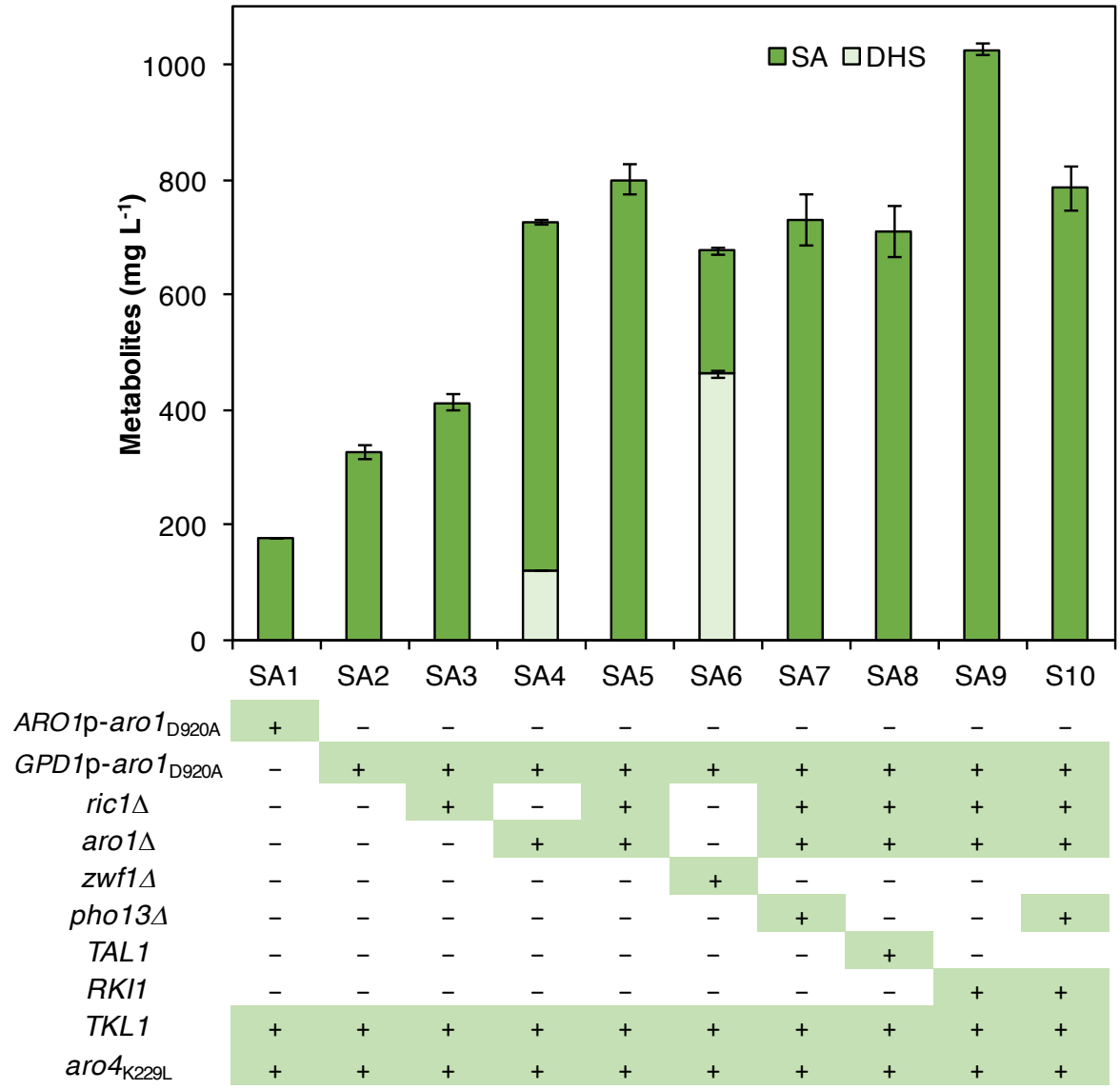


Figure. 4: Fermentation results from strains engineered to produce SA. Strain were grown in minimal media lacking histidine and supplemented with uracil and the three aromatic amino acids. Samples were collected at 72 h for HPLC analysis. The accumulation of DHS was only included for strain SA4 and SA6 since it was accumulated at a decent amount in these two strains. The variation is represented by the standard deviation from three biological replicates.

3.3.3. Discovery of novel metabolic interventions by OptForce

The OptForce algorithm intends to scope out non-intuitive interventions generally overlooked in manual investigation. For SA hyper-accumulation, OptForce first suggested two interventions consistent with what has been successfully implemented for producing aromatic amino acid pathway derivatives in the previous studies. These include, up-regulation of transketolase (*TKL1*) and 3-dehydroquinate synthase (DHQS, by replacing native *ARO1* with $ARO1_{D920A}$). In addition, pyruvate kinase (PYK) down-regulation by 2.9-fold was identified as one of the single interventions that could improve SA production. Down-regulation of PYK allows for the accumulation of precursor PEP (**Supplementary Table 2**). Note that while this intervention appeared to be detrimental towards up-regulating the aromatic amino acid pathway as observed previously (Gold et al., 2015), we believe that it was because this intervention was combined with the deletion of *ZWF1*, causing an NADPH deficiency in the organism as explained in Section 2.3.1.

OptForce has also suggested the rewiring of the PPP including up-regulation of transketolase (**Supplementary Table 2**), consistent with the manipulations introduced in the base strain (SA1) and previously reported (Curran et al., 2012; Suástegui et al., 2016b). In this work, however, the primary focus was on identifying interventions that, to the best of our knowledge, have not been previously linked to the production of aromatic amino acid pathway derivatives. In this regard, manipulations of four individual metabolic targets in the central carbon metabolism were identified, including: downregulation of the reactions 3-phophoglycerate kinase (*PGK1*) and glyceraldehyde-3-phosphate dehydrogenase (*TDH1*), knockout of phosphofructokinase 1 (*PFK1*), and overexpression of ribose-5-phosphate ketol-isomerase (*RKI1*) (**Figure 3a, b**, **Supplementary Table 2**). Overall, the intention of all these

interventions was to divert carbon flux from glycolysis towards the biosynthesis of the precursor E4P. Although all four interventions individually demonstrated an increase in the production of SA compared to the wild-type scenario based on *in silico* analysis, overexpression of *RKI1* seemed to be most feasible. Considering the type of fermentation regime studied here (aerobic and glucose-rich), deletion or knockdown of the genes *PGK1, TDH1,* and *PFK1*, could drastically reduce cell growth given their key roles in glycolysis. Hence, only the overexpression of *RKI1* was selected as the new intervention to be tested.

Ribose-5-phosphate ketol-isomerase catalyzes the interconversion of ribose-5-phosphate and ribulose-5-phosphate in the PPP (**Figure 3a**). According to OptForce, its overexpression could generate a higher flux into the aromatic amino acid pathway by maintaining a higher carbon pull into the non-oxidative PPP, *i.e.,* by directing the flux towards the formation of E4P and preventing it from recirculating back into glycolysis (**Figure 3a**). The *in silico* overexpression of *RKI1* resulted in the highest increase in SA compared to the other individual manipulations, achieving 83.45% of the theoretical yield. Furthermore, its combination with TKL1 overexpression, DHQS upregulation (enabled by *aro1*$_{D920A}$ overexpression), and SHKK knockout (enabled by *aro1* deletion) increased the yield to 0.55 g SA $g^{-1}_{glucose}$, representing 89.41% of the maximum theoretical yield based on *in silico* prediction (**Figure 3c**).

Experimentally, the overexpression of *RKI1* in strain SA5 (yielding strain SA9) resulted in a titer of 1,026.6 ± 51 mg L^{-1} SA (**Figure 4**). This strain constituted the highest-producing strain in this work with a yield of 51 mg $g^{-1}_{glucose}$, an increase of 4.82-fold compared to the initial strain. In a final attempt to increase the flux into the SA titers, we deleted *PHO13* on strain SA9 (yielding strain SA10). Although this strain improved the production by 11% compared to strain SA8, there was a decrease of about 23% compared to strain SA9 (**Figure 4**). This

indicates that overexpression of *RKI1* resulted in a consistent, beneficial manipulation for improving the flux in the aromatic amino acid pathway, but the higher transcription rate of *TAL1* resulting from *PHO13* deletion channeled the flux back into glycolysis, leading to a lower titer.

To further study the upper production limit of strain SA9, we performed fermentation in the media with increasing concentrations of carbon source. As demonstrated in **Figure 5**, elevating the sugar concentrations resulted in higher titers of SA. The first enhancement was observed from doubling the glucose concentration from 2% to 4%, in which the SA titer increased to 1.9 ± 0.1 g L^{-1} while maintaining a yield above 54 mg $g^{-1}_{glucose}$. The highest SA titer was achieved at 2.4 ± 0.051 g L^{-1}, with DHS accumulation at 193.2 ± 5.1 mg L^{-1}, in media containing 4% sucrose. The overall titer reached the highest production in *S. cerevisiae* among all aromatic amino acid pathway derivatives reported to date, although the glucose yield was compromised, dropping 42% compared to the fermentation in 2% glucose.

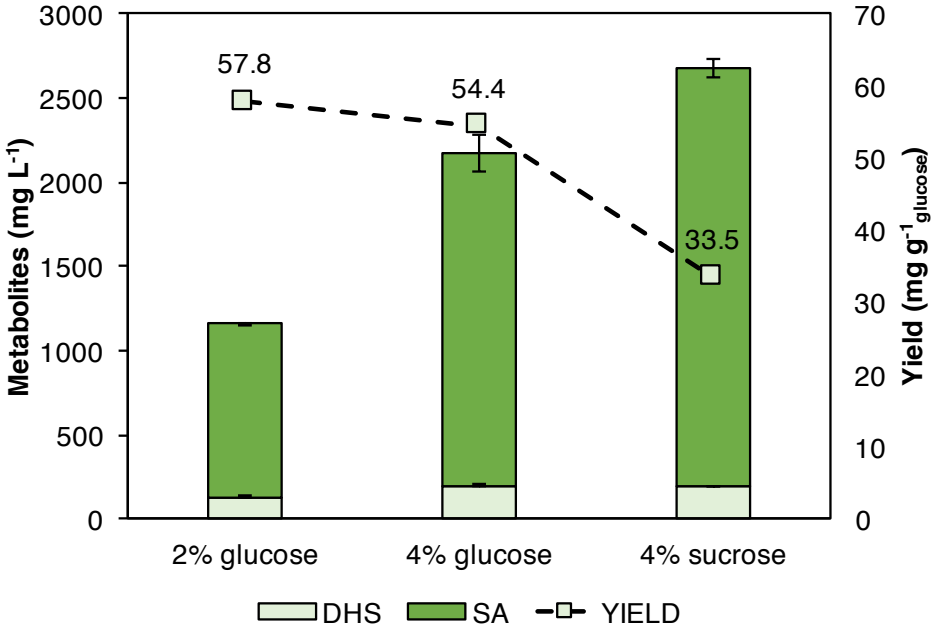


Figure 5: Variation of sugar concentration for SA fermentation. Strain SA9 was selected as the highest SA producer and tested for high glucose fermentations. Increasing the sugar concentration to 4% glucose and 4% sucrose increased the production of SA to 1.9 g L^{-1} and 2.4 g L^{-1}, respectively.

3.4. Case study: Muconic acid

To show the applicability of the manipulations that yielded the SA producing platform strain, we introduced the genetic elements to switch the production towards MA (**Table 5**). It is important to mention that unlike SA, MA is not an endogenous product in *S. cerevisiae*, hence the introduction of three heterologous genes from diverse microorganisms was required. According to previous studies, and also observed in this work, the gene *AroY* from the bacterium *Klebsiella pneumoniae* is the rate-limiting step in the pathway (Curran et al., 2012; Suástegui et al., 2016b). This gene encodes for protocatechuic acid (PCA) decarboxylase, and due to its oxygen sensitivity, it prevents an efficient conversion of PCA into catechol.

Therefore, to fully capture the effects of the metabolic manipulations, the comparison of titers will include the sum of PCA and MA from here on.

Table 5: Muconic acid producing strains.

Strain	Genotype	Plasmid	Parent
MA1	BY4741	pRS413- GPD1-*aro1*$_{D1409A,\text{D920A}}$ / TP1-*aro4*$_{\text{K229L}}$ / ADH1-*TKL*	BY4741
		pRS425- PYK1-*AROZ* / GPD1 – *HQD2* / TEF1 – *AROY*	
MA2	BY4741	pRS413-GPD1-*aro1*$_{D1409A,\text{D920A}}$ / TP1-*aro4*$_{\text{K229L}}$ /ADH1-*TKL* / PGK1-*RKI1*	BY4741
		pRS425- PYK1-*AROZ* / GPD1 – *HQD2* / TEF1 – *AROY*	
MA3	BY4741 *aro1*Δ *ric1*Δ	pRS413-GPD1-*aro1*$_{D1409A,\text{D920A}}$ / TP1-*aro4*$_{\text{K229L}}$ /ADH1-*TKL* / PGK1-*RKI1*	AA_003
		pRS425- PYK1-*AROZ* / GPD1 – *HQD2* / TEF1 – *AROY*	
MA4	BY4741 *aro1*Δ *ric1*Δ *zwf1*Δ	pRS413-GPD1-*aro1*$_{D1409A,\text{D920A}}$ / TP1-*aro4*$_{\text{K229L}}$ /ADH1-*TKL* / PGK1-*RKI1*	AA_006
		pRS425- PYK1-*AROZ* / GPD1 – *HQD2* / TEF1 – *AROY*	

Initially, the single-copy plasmid for DHS accumulation (pRS413-midAA_1) was co-transformed with the multiple-copy plasmid pRS425-MA to channel DHS into the production of MA. Compared to the platform strains for the production of SA, the MA strains harbored the mutant *aro1*$_{\text{D1409A-D920A}}$ to ensure complete blockage of the SA dehydrogenase subunit for accumulating DHS for MA synthesis (Suástegui et al., 2016b). This yielded the reference strain MA1, capable of accumulating 247.24 mg L^{-1} of MA and PCA. Overexpression of *RKI1*, yielding strain MA 2, increased the titer by 47%, achieving 363.97 mg L^{-1} (**Figure 6**). Deletion of *RIC1* and *ARO1* (strain MA3) increased the overall titer by another 3-fold, reaching 1105 mg L^{-1} (873 mg L^{-1} PCA plus 232 mg L^{-1} MA), a similar titer observed from strain SA9, indicating that the set of metabolic engineering strategies for the SA platform are translatable

to other products from the same module. The strategies implemented here were consistent with OptForce predictions, wherein different combinations of interventions involving transketolase (*TKL1/TKL2*), ribose-5-phosphate isomerase (*RKI1*), and 3-dehydroquinate synthase (*ARO1* mutant) led to the highest yield of MA (**Supplementary Table 2**).

In strain SA6 (**Figure 4**), DHS was accumulated due to an imbalance in reducing power caused by *ZWF1* deletion. Here for MA production, this manipulation was presumed to be beneficial since DHS is the precursor of MA pathway (also suggested by the OptForce analysis, **Figure 6** and **Supplementary Table 2**). Surprisingly, deletion of *ZWF1* (strain MA4) did not increase the production of PCA and MA in experiment; rather, it decreased the overall titer by 31%. This could be a result of removing the primary source of NADPH, in combination with a higher metabolic burden imposed on the MA producing strain, which carries the three-gene heterologous pathway in a multiple-copy plasmid along with multiple gene deletions. Despite that, it is noteworthy that OptForce was capable of distinguishing the redox requirements of SA and MA synthesis by suggesting *ZWF1* knockout for only MA production.

In order to test the upper limit of MA production, the glucose concentration was doubled, which led to an overall titer of 1598 mg L^{-1}, composed of 1267.1 ± 125.3 mg L^{-1} PCA and 330.9 mg L^{-1} ± 34.9 MA. Although this is the highest overall titer achieved, the yield was compromised, dropping from 55.3 to 40 mg $g^{-1}_{glucose}$ (**Figure 7**). The MA to PCA ratio was conserved at 0.26 when growing the strain in both 2% and 4% glucose.

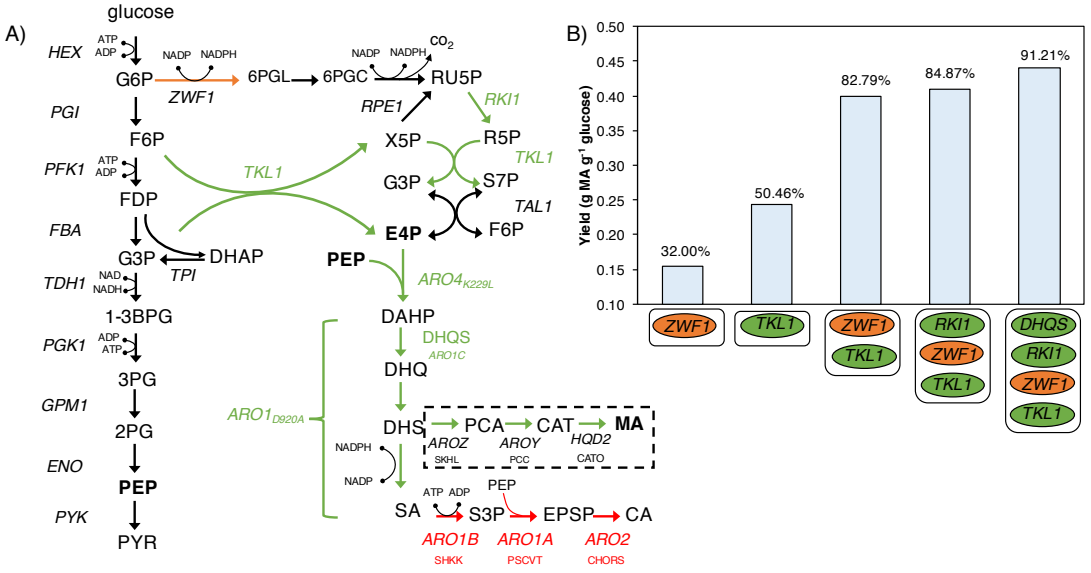


Figure 6: Metabolic interventions for the overproduction of MA identified through OptForce analysis. Green, red, and orange arrows (and circles) represent overexpression, downregulation and knockout of genes, respectively. A combination of *ZWF1* knockout, *TKL1*, *RKI1* and DHQS overexpression led to 0.441 g MA $g^{-1}_{glucose}$, which is equivalent to 91.21% of theoretical maximum yield.

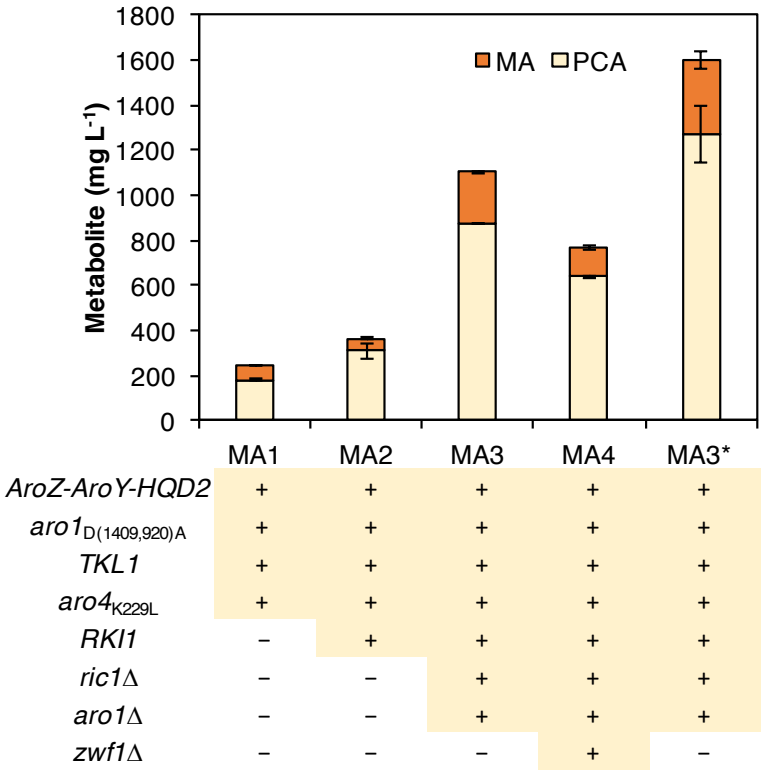


Figure 7: Fermentation results from strains engineered to produce MA. Strain were grown in minimal media lacking histidine and leucine and supplemented with uracil and the three aromatic amino acids. Samples were collected at 72 h for HPLC analysis. The variation is represented by the standard deviation from three biological replicates. The asterisk (*) indicates MA3 was grown in 4% glucose.

4. Discussion

Knowledge about the mechanism of transcriptional repression of a metabolic pathway can substantially facilitate the engineering efforts to construct microbial hosts with high production titers and yields. Thus far, only few reports have studied transcriptional regulation mechanisms in the aromatic amino acid pathway in the yeast *S. cerevisiae*. For example, it is known that the global TF Gcn4p can bind to the promoter sequences of several genes in the pathway in response to the levels of aromatic amino acids present in the culture media (Hinnebusch, 1988; Hinnebusch and Natarajan, 2002). Also, another TF Aro80p can activate the transcription of the genes *ARO9* and *ARO10* to initiate catabolism of the aromatic amino acids in the presence of a poor nitrogen source (Lee and Hahn, 2013). However, compared to other metabolic pathways in yeast, *e.g.,* glycolysis/glycogenesis and fatty acid metabolism, the knowledge of the aromatic amino acid pathway transcriptional repression remains understudied. As a result, only low production titers from this pathway have been achieved (Suástegui and Shao, 2016). Therefore, we decided to investigate, for the first time, which transcription factors could affect the carbon flux entrance, focusing our attention to the precursor module of the pathway.

Herein, we employed the web tool YEASTRACT (Teixeira et al., 2014) to create a functional link between TFs, and their effects on the production of aromatic amino acid pathway precursors. Combining it with a minimal, plasmid-based, SA-production platform corroborated the prediction capability of this database. By screening a small set of 21 TFs for knockout analysis, we already identified one regulator encoded by *RIC1*, which appears to negatively interact with the aromatic genes (*ARO2-4)*. The increases in titers caused by the deletion of *RIC1* were consistent throughout the constructed strains for both SA and MA production. This demonstrated the applicability of this tool to discover non-intuitive targets by

significantly narrowing the landscape of possible genetic interventions with the built-in ranking algorithms in YEASTRACT. Deletion of *RIC1* constituted the first level in enhancing the production of aromatic amino acid pathway-derived compounds. As such, it contributed to relieving the pathway from strong transcriptional repression, which, along with removal of feedback inhibition (*i.e.* expression of the mutant $aro4_{K229L}$), is key to unlocking higher carbon flux into the aromatic amino acid pathway.

According to the YEASTRACT analysis, there is evidence that the protein encoded by *RIC1* can also interact with the promoters of the genes *PHO3*, *PHO11*, and *PHO12*. Recently, a gene from the same family of phosphate metabolism regulators, *PHO13*, was found to act as a global regulator of the PPP (Kim et al., 2015). Deletion of this gene increased the transcription rate of several genes in the PPP, a characteristic that is beneficial for the assimilation of xylose. We initially hypothesized that such elevated transcription rates would also benefit the production of aromatic amino acid pathway derivatives. However, deletion of *PHO13* (strain SA7) decreased the SA titer, a phenotype that was also observed when the gene *TAL1* was overexpressed (strain SA8). This close relationship between the phosphate regulating family (*PHO*) and the biosynthesis of amino acid pathway derivatives led us to study the deletion of ten *PHO* genes and their effects on the production (**Supplementary Table 2**). The plasmid pRS413-midAA was transformed into all the *PHO* deletion strains and tested for the production of SA. Increase was not observed when comparing the titers to strains SA2. The lowest titer was 115 mg L^{-1} in strain BY4741 Δ*pho90* and the highest, 303.6 mg L^{-1}, was achieved in strain BY4741 Δ*pho88* (**Figure 8**). The products of both genes are involved in the transport of phosphate; hence the lower titers could be due to alterations in phosphate availability (Ghillebert et al., 2011; Yompakdee et al., 1996). Although these results indicate

that the carbon entry into the aromatic pathway is highly dependent on the PPP, and this pathway appears to be regulated by phosphate metabolism (*PHO* genes) (Kim et al., 2015), more in-depth studies are required to establish a link between these important metabolic modules.

As a second level of engineering, we deleted the gene *ARO1* to ensure full carbon capture for the products of interest (*i.e*. SA and MA). We observed that this intervention increased the titers higher than 50%; however, this was at the expense of adding slightly higher amounts of the three aromatic amino acids (50 mg L^{-1} of each amino acid).

The third level of engineering revolved around manipulating the structural genes to improve the precursor availability. This included the deletion of *ZWF1*, which, unlike the results from other reported work (Curran et al., 2012), did not benefit the production of either MA or SA (**Figure 4** and **Figure 7**). Deletion of this important gene in charge of supplying the cell with reducing power apparently affected the cells' fitness. Strains carrying this deletion required 2 to 3 more days to observe colonies on the selective plate, making this modification much less appealing for the construction of microbial cells for industrial-scale production. On the other hand, deleting *ZWF1* could be beneficial for the production of downstream metabolites by linking the NADPH deficiency to an enzyme such as prephenate dehydrogenase that regenerates NADPH (Gold et al., 2015). To achieve this, however, the available pool of NADPH should be able to sustain the reaction conversion of DHS to SA that requires NADPH as a cofactor at the first place. Overexpressing a heterologous NADPH recycling transhydrogenase could potentially help overcome this metabolic hurdle (Fiaux et al., 2003).

Within the same level of engineering, and as a result of implementing non-intuitive design predicted by the OptForce framework, we overexpressed the gene *RKI1* encoding the essential

enzyme ribose-5-phosphate isomerase. To our knowledge, this is the first time this gene has been overexpressed to enhance the flux into the aromatic amino acid pathway. Overexpressing *RKI1* led to the highest SA producing *S. cerevisiae* strain (SA9, **Figure 4**), due to an enhanced pull of carbon coming from the oxidative PPP and recycling of the product of the transketolase reaction (*i.e.* xylulose 5-phosphate, X5P) that is responsible for pulling flux from glycolysis back to PPP (**Figure 3a**). In combination with our previous manipulations, this overexpression enabled a titer of over 1 g L^{-1} and a yield of 51.3 mg $g^{-1}_{glucose}$, which is currently the highest recorded titer of all the compounds deriving from the aromatic amino acid pathway in *S. cerevisiae*. This final intervention further confirms that a higher pull into the PPP is required to enhance the pool of E4P.

Finally, it is important to remark that the new manipulations implemented here were easily transferred to extend the applicability of this platform to the production of MA. The MA-producing strain MA3 reached a total accumulation pool of MA and PCA that was 300% higher than that of the previously highest-producing strain reported under equivalent fermentation condition (Suástegui et al., 2016b). This evidences that the novel interventions implemented here significantly contributed towards redirecting carbon into the aromatic amino acid pathway. Given the economic relevance of this compound for the polymer industry, future work to improve strain MA3 should focus on improving the conversion of PCA to MA, which still remains as the bottleneck step. For example, a recently developed fast-screening platform for MA (Skjoedt et al., 2016) could be applied for directed evolution of the enzyme PCA decarboxylase to address its oxygen sensitivity. Finally, chromosomal integration of the production pathway could help eliminate the drawback arising from plasmid instability.

Overall, this work underlines the advantages of integrating multi-level information garnered from computational tools for unlocking the full capabilities of microorganisms. We have demonstrated that OptForce could suggest novel fruitful interventions that are consistent over multiple target products in the same metabolic branch as long as they obey the same constraints of energy and redox balance. We expect that the engineering framework implemented here will translate to the production of downstream metabolites of the SA pathway that have special nutraceutical and pharmaceutical values, such as the tyrosine-derived alkaloids, flavonoids, and stilbenoids with attention then switched to the heterologous expression of downstream proteins from plants. A new analysis of TFs targeting the genes located downstream of SA in combination with a minimal genetic platform that is modularly optimized should allow enhanced yields and titers of the highly valuable plant-sourced secondary aromatic metabolites.

5. Conclusion

Partitioning the long aromatic amino acid pathway in a modular fashion facilitated a multilevel approach for achieving high yields and titers. In this work, we demonstrate novel metabolic engineering strategies to enhance the production of aromatic amino acid pathway precursors. Among them, deletion of *RIC1* (based on the TF study) and overexpression of *RKI1* (based on the OptForce analysis) are two novel interventions that are crucial to enable the production of aromatic amino acid pathway precursors at titers higher than 1 g L^{-1}, with the highest reported yields thus far in simple batch fermentation. This is a significant milestone considering the previous titers of the vast majority of the compounds derived from this resourceful pathway in the batch fermentation were constrained at low levels (Suástegui and

Shao, 2016). Future work will focus on transferring this engineering framework towards the production of the higher-value chemicals downstream of the precursor module.

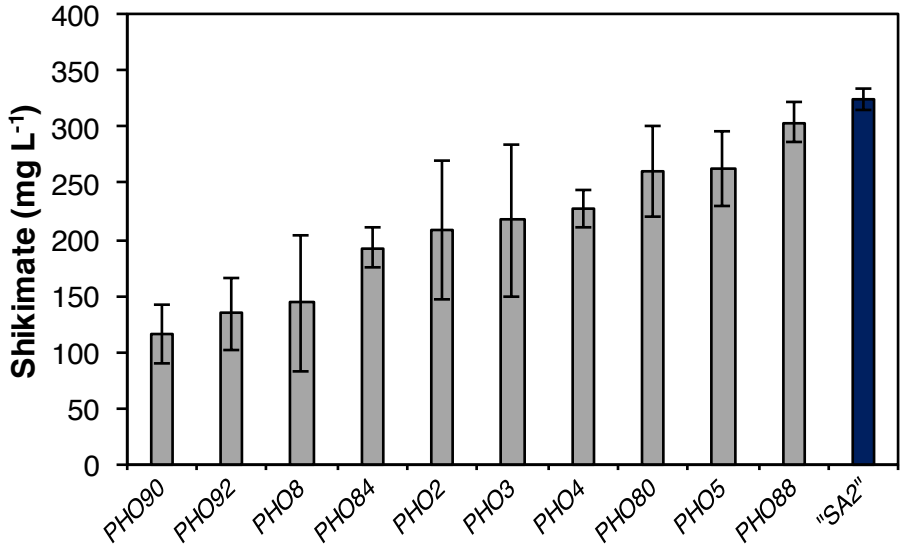


Figure 8: Fermentation results from *PHO* KO strains

.

6. References

Amberg, D. C., Burke, D. J., Strathern, J. N., 2006. "Quick and dirty" plasmid transformation of yeast colonies. CSH Protoc. 2006.

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., Zhao, H., 2014. Homology-Integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. ACS Synth. Biol. 4**,** 585-94.

Bensen, E. S., Yeung, B. G., Payne, G. S., 2001. Ric1p and the Ypt6p GTPase function in a common pathway required for localization of trans-Golgi network membrane proteins. Mol. Biol. Cell. 12**,** 13-26.

Biggs, B. W., De Paepe, B., Santos, C. N., De Mey, M., Kumaran Ajikumar, P., 2014. Multivariate modular metabolic engineering for pathway and strain optimization. Curr. Opin. Biotechnol. 29**,** 156-62.

Brochado, A. R., Matos, C., Møller, B. L., Hansen, J., Mortensen, U. H., Patil, K. R., 2010. Improved vanillin production in baker's yeast through in silico design. Microb. Cell Fact. 9**,** 84-84.

Brown, J. F., Dawes, I. W., 1990. Regulation of chorismate mutase in *Saccharomyces cerevisiae*. Mol. Genet. Genomics. 220**,** 283-8.

Chowdhury, R., Chowdhury, A., Maranas, C. D., 2015. Using Gene Essentiality and Synthetic Lethality Information to Correct Yeast and CHO Cell Genome-Scale Models. Metabolites. 5**,** 536-70.

Chung, C. T., Niemela, S. L., Miller, R. H., 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. U. S. A. 86**,** 2172-5.

Curran, K. A., Leavitt, J. M., Karim, A. S., Alper, H. S., 2012. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab. Eng. 15**,** 55-66.

de Boer, M., Nielsen, P. S., Bebelman, J. P., Heerikhuizen, H., Andersen, H. A., Planta, R. J., 2000. Stp1p, Stp2p and Abf1p are involved in regulation of expression of the amino acid transporter gene *BAP3* of *Saccharomyces cerevisiae*. Nucleic Acids Res. 28**,** 974-81.

Engler, C., Gruetzner, R., Kandzia, R., Marillonnet, S., 2009. Golden Gate Shuffling: A one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One. 4**,** e5553.

Engler, C., Kandzia, R., Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One. 3**,** e3647.

Feng, X., Zhao, H., 2013a. Investigating glucose and xylose metabolism in *Saccharomyces cerevisiae* and *Scheffersomyces stipitis* via ^{13}C metabolic flux analysis. AIChE J. 59**,** 3195-3202.

Feng, X., Zhao, H., 2013b. Investigating xylose metabolism in recombinant *Saccharomyces cerevisiae* via ^{13}C metabolic flux analysis. Microb. Cell Fact. 12**,** 114.

Fiaux, J., Cakar, Z. P., Sonderegger, M., Wuthrich, K., Szyperski, T., Sauer, U., 2003. Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. Eukaryot. Cell. 2**,** 170-80.

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., Smolke, C. D., 2015. Complete biosynthesis of opioids in yeast. Science. 349**,** 1095-100.

Gao, M., Cao, M., Suastegui, M., Walker, J. A., Rodriguez-Quiroz, N., Wu, Y., Tribby, D., Okerlund, A., Stanley, L. M., Shanks, J. V., Shao, Z., 2016. Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. ACS Synth. Biol.

Ghillebert, R., Swinnen, E., De Snijder, P., Smets, B., Winderickx, J., 2011. Differential roles for the low-affinity phosphate transporters Pho87 and Pho90 in *Saccharomyces cerevisiae*. Biochem. J. 434**,** 243-51.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., III, Smith, H. O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods. 6**,** 343.

Gold, N. D., Gowen, C. M., Lussier, F.-X., Cautha, S. C., Mahadevan, R., Martin, V. J. J., 2015. Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics. Microb. Cell Fact. 14**,** 73.

Hinnebusch, A. G., 1988. Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. Microbiol Rev. 52**,** 248-73.

Hinnebusch, A. G., Natarajan, K., 2002. Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. Eukaryot. Cell. 1**,** 22.

Jiang, H., Wood, K. V., Morgan, J. A., 2005. Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 71**,** 2962-9.

Kim, S. R., Xu, H., Lesmana, A., Kuzmanovic, U., Au, M., Florencia, C., Oh, E. J., Zhang, G., Kim, K. H., Jin, Y. S., 2015. Deletion of PHO13, encoding haloacid dehalogenase type IIA phosphatase, results in upregulation of the pentose phosphate pathway in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 81**,** 1601-9.

Koopman, F., Beekwilder, J., Crimi, B., Van Houwelingen, A., Hall Robert, D., Bosch, D., Van Maris Antonius, J., Pronk Jack, T., Daran, J.-M., 2012. *De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. Microb. Cell Fact. 11**,** 155.

Lee, K., Hahn, J. S., 2013. Interplay of Aro80 and GATA activators in regulation of genes for catabolism of aromatic amino acids in *Saccharomyces cerevisiae*. Mol. Microbiol. 88**,** 1120-34.

Li, M., Kildegaard, K. R., Chen, Y., Rodriguez, A., Borodina, I., Nielsen, J., 2015. *De novo* production of resveratrol from glucose or ethanol by engineered *Saccharomyces cerevisiae*. Metab. Eng. 32**,** 1-11.

Luttik, M. A. H., Vuralhan, Z., Suir, E., Braus, G. H., Pronk, J. T., Daran, J. M., 2008. Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: Quantification of metabolic impact. Metab. Eng. 10**,** 141-153.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., Valen, E., 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. 42**,** W401-7.

Qian, W., Ma, D., Xiao, C., Wang, Z., Zhang, J., 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep. 2**,** 1399-410.

Ranganathan, S., Suthers, P. F., Maranas, C. D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6**,** e1000744.

Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., Nielsen, J., 2015. Establishment of a yeast platform strain for production of *p*-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. Metab. Eng. 31**,** 181-188.

Shao, Z., Luo, Y., Zhao, H., 2012. DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*. Methods Mol. Biol. 898**,** 251-62.

Shao, Z., Zhao, H., 2013. Construction and engineering of large biochemical pathways via DNA assembler. Methods Mol. Biol. 1073**,** 85-106.

Shao, Z., Zhao, H., 2014. Manipulating natural product biosynthetic pathways via DNA assembler. Curr. Protoc. Chem. Biol. 6**,** 65-100.

Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J., Siedler, S., Borodina, I., Jensen, M. K., Keasling, J. D., 2016. Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. Nat. Chem. Biol. 12**,** 951-958.

Suástegui, M., Guo, W., Feng, X., Shao, Z., 2016a. Investigating strain dependency in the production of aromatic compounds in *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 113.

Suástegui, M., Matthiesen, J. E., Carraher, J. M., Hernandez, N., Rodriguez Quiroz, N., Okerlund, A., Cochran, E. W., Shao, Z., Tessonnier, J. P., 2016b. Combining metabolic engineering and electrocatalysis: application to the production of polyamides from sugar. Angew. Chem., Int. Ed. 55**,** 2368-73.

Suástegui, M., Shao, Z., 2016. Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. J. Ind. Microbiol. Biotechnol. 43**,** 1611-1624.

Teixeira, M. C., Monteiro, P. T., Guerreiro, J. F., Goncalves, J. P., Mira, N. P., dos Santos, S. C., Cabrito, T. R., Palma, M., Costa, C., Francisco, A. P., Madeira, S. C., Oliveira, A. L., Freitas, A. T., Sa-Correia, I., 2014. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. Nucleic Acids Res. 42**,** D161-6.

Trantas, E., Panopoulos, N., Ververidis, F., 2009. Metabolic engineering of the complete pathway leading to heterologous biosynthesis of various flavonoids and stilbenoids in *Saccharomyces cerevisiae*. Metab. Eng. 11**,** 355-66.

Tsolas, O., Horecker, B. L., 1973. Transaldolase: a model for studies of isoenzymes and half-site enzymes. Mol. Cell. Biochem. 1**,** 3-13.

Vilela Lde, F., de Araujo, V. P., Paredes Rde, S., Bon, E. P., Torres, F. A., Neves, B. C., Eleutherio, E. C., 2015. Enhanced xylose fermentation and ethanol production by engineered *Saccharomyces cerevisiae* strain. AMB Express. 5**,** 16.

Xu, P., Gu, Q., Wang, W., Wong, L., Bower, A. G., Collins, C. H., Koffas, M. A., 2013. Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. Nat. Commun. 4**,** 1409.

Yan, Y., Kohli, A., Koffas, M. A., 2005. Biosynthesis of natural flavanones in Saccharomyces cerevisiae. Appl. Environ. Microbiol. 71**,** 5610-3.

Yompakdee, C., Ogawa, N., Harashima, S., Oshima, Y., 1996. A putative membrane protein, Pho88p, involved in inorganic phosphate transport in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 251**,** 580-90.

Zhang, J., ten Pierick, A., van Rossum, H. M., Seifar, R. M., Ras, C., Daran, J. M., Heijnen, J. J., Wahl, S. A., 2015. Determination of the cytosolic NADPH/NADP ratio in *Saccharomyces cerevisiae* using shikimate dehydrogenase as sensor reaction. Sci. Rep. 5**,** 12846.

Zomorrodi, A. R., Maranas, C. D., 2010. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. BMS Syst. Biol. 4**,** 178.

7. Supplementary Information

7.1. Supplementary Methods

7.1.1. Identification of pathway interventions using OptForce

In all simulations, the maximum glucose and oxygen uptake rates were set to 100 mmol g DW^{-1} h^{-1} and 200 mmol g DW^{-1} h^{-1}, respectively. The regulation on the tricarboxylic acid cycle (TCA) cycle activity under aerobic glucose conditions (*i.e.* the Crabtree effect) was originally imposed in the *i*AZ900 model by limiting the oxygen uptake rate (Zomorrodi and Maranas, 2010). This is replaced in this simulation by directly imposing an upper bound on the mitochondrial cytochrome c oxidase (CYOOm) reaction flux to 20 mmol g DW^{-1} h^{-1}. This modification, while simultaneously preserving the phenotypes observed due to the Crabtree effect in the model, did not restrict additional oxygen to be consumed in the MA pathway. Under minimal media condition, the maximum yields were 2.89 h^{-1} for biomass, 64 mmol g DW^{-1} h^{-1} for SA, and 62.13 mmol g DW^{-1} h^{-1} for MA.

In brief, the procedure for identifying the reaction-level interventions is described as follows. See detailed description of the procedure in our previous work (Ranganathan et al., 2010).

Step 1: MFA data for wild type *S. cerevisiae* (Jacqueline Shanks, Iowa State University, personal communication) was used to constrain the flux ranges for reactions in the central metabolism. Flux variability analysis (FVA) is performed to identify the wild type strain flux ranges ($[v_j^{ref,L}, v_j^{ref,U}]$).

Step 2: The flux ranges for overproducing strain ($[v_j^{OS,L}, v_j^{OS,U}]$) are identified by performing FVA on the metabolic network that is subjected to the desired overproduction target (90% of v_{MA}^{max} or v_{SA}^{max}) and at least 10% of theoretical biomass yield ($v_{biomass}^{max}$).

Step 3: By superimposing the flux ranges for the wild type and the overproducing strain, the set of reactions that must be up-regulated ($MUST^U$), down-regulated ($MUST^L$) and knockout ($MUST^X$) are then determined. Second order MUST sets (*i.e.* $MUST^{UU}$, $MUST^{UL}$ and $MUST^{LL}$) were also identified according to previous work (Ranganathan et al., 2010).

Step 4: The minimal set of interventions (FORCE sets) that guarantees the yield of the target product under worst-case scenario was then selected from the MUST sets. To this end, a bi-level mixed-integer optimization problem was formulated such that the objectives of the outer and the inner problem are the maximization and the minimization of the target product yield, respectively (Chowdhury et al., 2015). The lower bound of the biomass formation is set to 10% of $v_{biomass}^{max}$ to ensure viability. The flux range for each reaction excluding biomass formation and nutrients uptake is restricted to $[\min\{v_j^{ref,L}, v_j^{OS,L}\}, \max\{v_j^{ref,U}, v_j^{OS,U}\}]$. Note that the minimum guaranteed production flux and yield for each of the mutants were calculated under the same condition as Step 4. All optimization problems were solved in GAMS 24.4.1 with CPLEX Solver v.12.6.1.

7.2. Supplementary Tables

Supplementary Table 1: List gRNA used for gene knockouts via the CRISPR system. The strikethrough represents the 8 bp deletion which includes the protospacer adjacent motif (PAM) in bold.

Target gene	gRNA	Donor DNA
ARO1	gttatggtagcgatcggtggtgg	tagaagattatcttttagtggaaggatgtactcgtgatacggttatggtagcga tcggtggtggtgttattggtgacatgattgggttcgttgcatctacatttatga
PHO13	gtatgggtatttggagaaagcgg	gactttctgaaattgcagcctggcaaagataaggtatgggtatttggagaaa gcggtattggtgaagaattgaaactaatggggtacgaatctctaggaggtg ccgat
ZWF1	gtccaaattgtccatggaggagg	tgatccatctaccaagatcttcggttatgcccggtccaaattgtccatggagg aggaactgaagtcccgtgtcctaccccacttgaaaaaacctcacggtgaag ccga

Supplementary Table 2: Metabolic interventions predicted by OptForce for (a) shikimic acid (SA) and (b) muconic acid (MA) overproduction. The theoretical maximum and minimum guaranteed fluxes are calculated under the condition specified in Supplementary Methods. Although multiple isozymes may be associated with each reaction only the major genes described in this study are shown in the gene-level interventions. Reaction abbreviations and the chemical equations written in the net flux direction are as followed:

(i) DHQS, 3-dehydroquinate synthase (2dda7p => 3dhq + pi);

(ii) G6PDH2, glucose-6-phosphate dehydrogenase (g6p + nadp => 6pgl + h + nadph);

(iii) GAPD, glyceraldehyde-3-phosphate dehydrogenase (g3p + nad + pi <=> 13dpg + h + nadh);

(iv) PFK, phosphofructokinase (atp + f6p => adp + fdp + h);

(v) PGK, phosphoglycerate kinase (3pg + atp <=> 13dpg + adp);

(vi) PYK, pyruvate kinase (adp + h + pep => atp + pyr);

(vii) RPI, ribose-5-phosphate isomerase (ru5p <=> r5p);

(viii) SHKK, shikimate kinase (atp + skm => adp + h + skm5p);

(ix) SKHL, 3-dehydroshikimate dehydratase (3dhsk <=> h2o + pca);

(x) TKT1, transketolase 1 (r5p + xu5p <=> g3p + s7p);

(xi) TKT2, transketolase 2 (f6p + g3p <=> e4p + xu5p).

Supplementary Table 2 continued

Metabolic interventions (Reaction level)	Metabolic interventions (Gene level)	Minimum guaranteed flux (mmol gDW^{-1} h^{-1})	Yield ($g_{product}$ $g^{-1}_{glucose}$)	% Theoretical maximum
(a) Shikimic acid				
↓PGK	↓*PGK1*	51.59	0.499	80.61%
↓GAPD	↓*TDH1*	51.59	0.499	80.61%
ΔPFK	Δ*PFK1*	51.85	0.501	81.02%
↑RPI	↑*RKI1*	53.41	0.516	83.45%
↑TKT1	↑*TKL1*	54.71	0.526	85.48%
↓PYK	↓CDC19	55.06	0.529	86.03%
↑TKT2	↑*TKL1*	55.52	0.537	86.75%
↑TKT1 ↑TKT2	↑*TKL1*	56.19	0.543	87.80%
↑TKT1 ↑TKT2 ↑RPI	↑*TKL1* ↑*RKI1*	56.19	0.543	87.80%
↑TKT1 ↑TKT2 ↓SHKK	↑*TKL1* ↓*ARO1*	56.49	0.546	88.27%
↑TKT1 ↑TKT2 ↑DHQS	↑*TKL1* ↑$ARO1_{D920A}$	57.13	0.552	89.27%
↑TKT1 ↑TKT2 ↑DHQS ↓SHKK ↑RPI	↑*TKL1* ↑$ARO1_{D920A}$ ↓*ARO1* ↑*RKI1*	57.22	0.553	89.41%
(b) Muconic acid				
ΔG6PDH2	Δ*ZWF1*	19.88	0.155	32.00%
↑RPI	↑*RKI1*	25.07	0.195	40.35%
↑TKT1	↑*TKL1*	31.35	0.244	50.46%
↑TKT1 ΔG6PDH2	↑*TKL1* Δ*ZWF1*	51.44	0.4	82.79%
↑TKT2	↑*TKL1*	51.63	0.401	83.10%
↑TKT1 ↑RPI ΔG6PDH2	↑*TKL1* ↑*RKI1* Δ*ZWF1*	52.73	0.41	84.87%
↑TKT2 ↑TKT1	↑*TKL1*	53.24	0.414	85.69%
↑TKT1 ↑RPI ΔG6PDH2 ↑SKHL	↑*TKL1* ↑*RKI1* Δ*ZWF1* ↑*AROZ*	55.92	0.435	90.00%
↑TKT2 ↑TKT1 ↑SKHL	↑*TKL1* ↑*AROZ*	55.92	0.435	90.00%

Supplementary Table 2 continued

↑TKT1 ↑DHQS ↑RPI	↑*TKL1* ↑$ARO1_{D920A}$ ↑*RKI1*	56.67	0.441	91.21%
↑TKT1 ↑TKT2 ↑DHQS ↑RPI	↑*TKL1* ↑$ARO1_{D920A}$ ↑*RKI1*	56.67	0.441	91.21%
↑TKT1 ↑TKT2 ↑DHQS ↑RPI ΔG6PDH2	↑*TKL1* ↑$ARO1_{D920A}$ ↑*RKI1* Δ*ZWF1*	56.67	0.441	91.21%
↑TKT1 ↑DHQS ↑RPI ΔG6PDH2	↑*TKL1* ↑$ARO1_{D920A}$ ↑*RKI1* Δ*ZWF1*	56.67	0.441	91.21%

CHAPTER 6

NEW STRATEGIES, NEW PRODUCTS AND NEW SPECIES

1. Introduction

As studied in the previous chapters, the carbon entrance into the aromatic amino acid pathway depends on the availability of two main precursors. A fine balance between erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) is expected to aid the production of metabolites deriving from the aromatic pathway. E4P is a product from the pentose phosphate pathway and various successful engineering strategies have been devised to enhance its accumulation (Suástegui and Shao, 2016). In contrast, strategies to increase the availability of phosphoenolpyruvate, the second precursor into the pathway, have not been established thus far in *Saccharomyces cerevisiae.* In prokaryotic microorganisms like *Escherichia coli*, a few manipulations have linked a higher production of shikimic acid pathway derivatives with a higher availability of PEP. For instance, the phosphotransferase system (PTS) in charge of transporting glucose into the cytoplasm can be knocked out, which reduces the utilization of PEP as the phosphate donor. Another successful strategy consists of recirculating pyruvate to PEP by means of overexpressing the phosphoenolpyruvate synthase, *ppsA* (Patnaik and Liao, 1994).

Intrinsic differences between prokaryotes and eukaryotes prevent the direct transfer of metabolic engineering strategies. For example, the transportation of glucose in *S. cerevisiae* does not depend on the PTS system. Furthermore, the presence of organelles results in the compartmentalization of metabolites, which reduces the availability of the desired precursors in the cytoplasm and dilutes the possibility of reaching higher product titers. PEP is a

cytoplasmic metabolite, meaning it could be readily available for condensation with E4P for the production of aromatic compounds. However, our metabolic flux analyses showed that the rates of PEP conversion to pyruvate are at least one order of magnitude higher than the E4P flux in the pentose phosphate pathway (Suástegui et al., 2016a). This is expected because pyruvate has considerably more metabolic fates in the cell, hence it is in higher demand for production of essential metabolites in aerobic respiration, fatty acid production, and production of other amino acids.

Therefore, this section is dedicated toward finding novel alternatives to rewire the availability of PEP to allow increases in aromatic titers while maintaining a good metabolic balance for efficient biomass formation. Moreover, the production of molecules from the downstream module is presented and the non-conventional yeast is studied for the production of coumaric acid.

2. New Strategies for Enhancing the Production of Aromatic Amino Acid Derivatives

Slowing down the conversion of PEP to pyruvate is desired to reduce the competition between the aromatic amino acid production and other essential pathways. The most straightforward manipulations revolve around deleting or downregulating reactions that consume PEP or pyruvate. However, a detrimental effect is expected as it is likely that *S. cerevisiae* will lose its capability of growing in fermentable sugars. Some examples for this rationale include deleting the pyruvate kinase gene *pyk1* (CDC19), which converts PEP to pyruvate, and deleting genes from the pyruvate decarboxylase family *PDC* to reduce the loss of carbon to ethanol and acetate (Suástegui et al., 2016b). However, neither of these techniques have resulted in positive outcomes which evidence that more dynamic strategies need to be implemented that 1)

consider the compartmentalization of metabolites, 2) overcome strong regulatory mechanisms, and 3) do not rely on knockout of essential genes.

2.1. Direct recirculation of pyruvate to PEP

There are several potential alternatives consisting mainly in gene overexpressions to recirculate pyruvate back to PEP. In Chapter 2, a strategy consisting on overexpressing the endogenous *S. cerevisiae* genes namely, pyruvate carboxylase *PYC1* and phosphoenolpyruvate carboxykinase *PCK1*. This rationale was established with the goal of enforcing the conversion of pyruvate to oxaloacetate (OAA), which would then be converted to PEP. This strategy, however, did not consider key regulatory points that control the degradation of gluconeogenic proteins (PCK1p), and also did not account for the transport of pyruvate to other organelles such as the mitochondria. Hence, a new strategy for pyruvate recirculation will explore the overexpression of heterologous genes, namely *ppsA* from *E. coli*, and the pyruvate, orthophosphate dikinase (PPDK) from the plant *Arabidopsis thaliana.* One of the advantages of overexpressing heterologous genes is that posttranslational modifications that inactivate proteins (phosphorylation, ubiquitylation) can be avoided.

To ensure a higher pyruvate availability in the cytosol to serve as the substrate for the recirculation enzymes mentioned above, the main carriers that transport pyruvate into the mitochondria need to be knocked out. This strategy has been previously established in *S. cerevisiae* to increase the cytosolic pools of pyruvate to elevate the production polyketides (Cardenas and Da Silva, 2016). In this study, two mitochondrial, outer membrane transporters (*POR1* and *POR2*) and two mitochondrial pyruvate inner membrane carriers (*MPC1* and *MPC2*) were deleted. Single-gene knockouts revealed that *POR2* and *MPC2* were good

candidates to increase the pool of pyruvate in cytoplasm resulting in at least a 3-fold improvement in triacetic acid lactone (TAL) production. The combination of these further increased the titers of TAL from ~ 300 mg L^{-1} to almost 1.0 g L^{-1}.

By combining the overexpression of the pyruvate recirculation genes in strains with the deletion of pyruvate mitochondrial transporters enhances in the availability of PEP to be channeled into the aromatic amino acid pathway are expected.

2.2 Enforcing gluconeogenesis for PEP accumulation

Gluconeogenesis is a metabolic pathway whose activation, in *S. cerevisiae* and other organisms, is triggered by a diauxic shift. This is, as glucose is depleted, the glycolytic genes undergo downregulation, while the gluconeogenic ones become activated with the goal of regenerating glucose from intracellular metabolites. Because the gluconeogenic pathway is highly ATP-consuming, very strict regulations at all metabolic levels protect the cell from unnecessary energy expenditures in the presence of high concentrations of glucose.

Evidence of the diauxic shift (initiation of gluconeogenesis) and its impact on the production of aromatic amino acid derivatives can be observed from the fermentation profiles of shikimic acid (See supplementary information in Chapter 4). At 48 h of fermentation, the glucose levels have been completely depleted, and it is at this point when the accumulation of shikimic acid significantly increases. At 72 h the maximum titer is usually achieved, indicating that during this time the levels of PEP increase, hence allowing the DAHP synthase to carry out the first committed step in the shikimic acid pathway more efficiently.

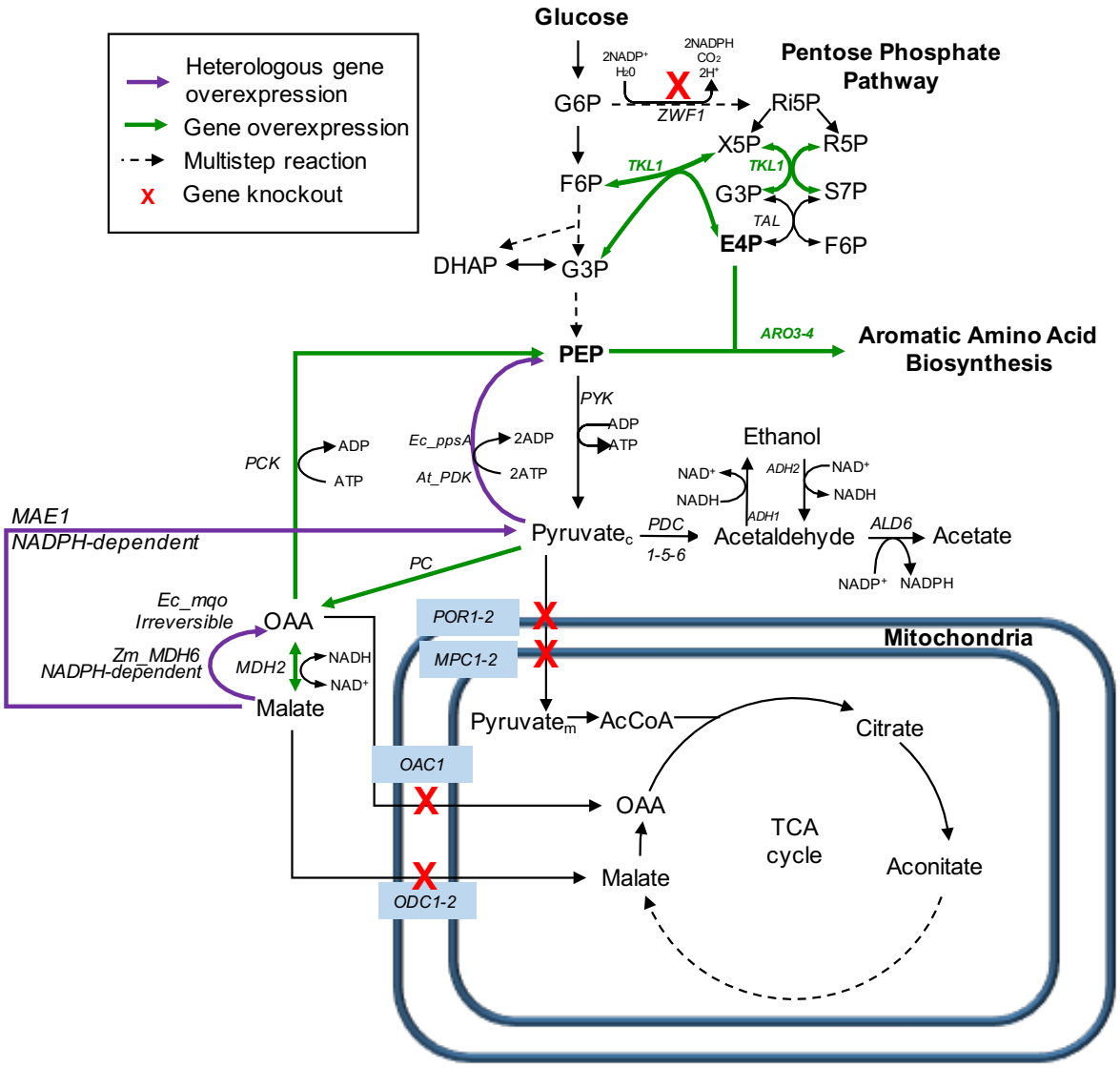


Figure 1: New strategies for improving the production of aromatic amino acid derivatives. The engineering rationale is based on the recirculation of pyruvate to PEP by a direct catalytic step requiring the overexpression of heterologous genes such as *Ec_ppsA*, or *At_PDK*. Another strategy includes the overexpression of endogenous gluconeogenic genes to accumulate PEP through the *PC-PCK* cycle or by MDH genes. This strategy also considers the deletion of transporter proteins to prevent the loss of carbon into mitochondria.

Therefore, a new metabolic engineering strategy will be implemented with the goal of enforcing the activation of gluconeogenic genes even in the presence of high glucose concentrations.

In *S. cerevisiae,* gluconeogenesis starts from the conversion of cytosolic malate to pyruvate through an $NADH^{+}$ dependent malic acid enzyme (pyruvic-malic enzyme, MAE1p) or through the conversion of malate to OAA through the cytosolic malic acid dehydrogenase enzyme (MDH2p) (**Figure 1**). OAA is further converted to PEP through the PCK1 enzyme. Seven more steps proceed this reaction to obtain glucose-6-phosphate, however, because the target metabolite for the production of aromatic is PEP, the attention will be focused on the steps that convert malate to PEP. Similar to pyruvate, malate can also be transported into the mitochondria through the action of the specific carrier *OAC1*. This is an inner membrane transporter that imports the cytosolic OAA produced from pyruvate by the enzyme pyruvate carboxylase (Palmieri et al., 1999). Therefore, the deletion of this gene could be a potential strategy to increase high malate pools for the initiation of gluconeogenesis (**Figure 1**).

Several regulatory mechanisms affect the gluconeogenic pathway at the transcriptional, post-transcriptional, and post-translational levels. For instance, it is known that strength of the *PCK1* promoter is detectable only during the stationary phase of cell growth (Reider Apel et al., 2017). This is caused by the action of several carbon-responsive elements such as Mig1p, which can repress the transcription of gluconeogenic transcriptional activators through binding to specific promoter regions (Schuller, 2003). Moreover, the mRNA of key gluconeogenic genes are destabilized by the action of the Reg1p protein which is associated with the Mig1p glucose repression pathway (Yin et al., 2000). These drawbacks can be easily overcome by

overexpressing the two genes of interest, namely *MDH2* and *PCK1*, with strong constitutive promoters which are not subjected to glucose repression. However, given the potential impact on cellular metabolism, it would be preferable to span a panel of promoters with different strengths to ensure an appropriate expression balance.

Post-translational modifications (PTMs) serve as regulatory commands to control the stability, localization, and activity of proteins in the cell. One or more amino acid residues in a protein can be modified by the addition of a chemical group (*e.g.* phosphorylation) to induce a specific switch-like state of activation of deactivation. Two of the most common post-translational modifications in eukaryotic systems are phosphorylation and ubiquitylation. The former is usually associated with enzymatic activity, whereas the latter serves as a tag for protein degradation. PTMs are commonly orchestrated by a network of genes and its study has been determinant in understating major hierarchical regulatory mechanisms of key metabolic pathways. Hence, it is not unexpected that PTMs play a crucial role in the regulation of highly energy-demanding pathways such as gluconeogenesis and the biosynthesis of aromatic amino acid. Several studies have revealed that, indeed, the gluconeogenic genes undergo major PTMs as an additional control level to regulate their cellular activity and abundance. For example, the cytosolic MDH protein (MDH2p), contains two phosphorylation residues at the proline 1 (P1) residue and at the threonine 6 (T6) residue. Through biochemical characterizations, it has been demonstrated that substituting P1 to serine (P1S), or removing the first 12 residues of the protein can reduce its degradation when cells are shifted from non-fermentable carbon sources (acetate) to growth on glucose (Hung et al., 2004; Minard and McAlister-Henn, 1992). Through global proteomic studies, four lysine residues (K42, K180, K254, and K259) subjected to ubiquitylation have also been identified in MDH2p (Swaney et al., 2013). Whether

ubiquitylation of these residues is a product of cross-talk between the phosphorylated residues (*i.e.* phosphodegrons that induce ubiquitylation) remains unknown, however, it would be interesting to study these interactions to discern whether a higher activity of this enzyme can be achieved in the presence of glucose. Therefore, different mutant versions of the gene *MDH2* will be constructed and incorporated into the new genetic platform to study the recirculation of malate to OAA.

To finalize the partial reconstruction of gluconeogenesis, the phosphoenolpyruvate carboxykinase gene, *PCK1*, will also be overexpressed, in order to ensure the conversion of OAA to PEP. Thus far there is no evidence of PCK1p undergoing PTMs such as phosphorylation or ubiquitylation. However, high sequence similarity with the protein ARO4p (50 % identity in the first 12 amino acids), which is known to undergo phosphorylation was observed. Therefore, new mutant versions of PCK1p with amino acid substitutions to avoid phosphorylation will be constructed and overexpressed to complete the production of PEP in the presence of glucose.

Figure 2: Sequence alignment of the first 24 amino acid residues of the *S. cerevisiae* PCK1p and ARO4p proteins. The serine residues in blue in ARO4p represent the recognized phosphorylation sites (Albuquerque et al., 2008). The sequence similarity of these proteins indicates that the Serine residues in PCK1 could be substituted to avoid phosphorylation.

3. Materials and Methods

3.1. Strains and plasmid construction

The strains constructed are listed in **Table 1**. Briefly, we used the haploid strain of *S. cerevisiae* BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and the *oac1*Δ BY4741 knockout strain to study the enforcement of gluconeogenesis through the overexpression of *MDH1* and the phosphorylation-free $MDH1_{P1S}$ variant. In this study, we also constructed a strain carrying a single cassette integration of the minimal SA pathway, described in Chapter 5.

The plasmid-free strains were propagated in YPAD media (1% yeast extract, 2% peptone, 0.01% adenine, and 2% dextrose). The plasmid harboring the SA pathway (SA_high, chapter 5) and the plasmid harboring the *MDH2* variants were co-transformed into the corresponding *S. cerevisiae strains* following the quick and dirty protocol (Amberg et.al., 2016). After plasmid transformation, yeast strains were plated in Synthetic Complete solid media lacking histidine and leucine and supplemented, to a final concentration 50 mg L^{-1}, with each of the three aromatic amino acids. All plasmids constructed in this study (**Table 1**) were derived from the pRS shuttle vector series (New England BioLabs, Ipswich, MA). To assemble the genetic cassettes, the DNA Assembler technique (Shao et al., 2012; Shao and Zhao, 2013) was implemented.

Table 1: The constructed plasmid and strains.

Strain	Plasmid	Expression cassette
BY4741 **Control**	pRS415	Empty
	pRS413-SA_high	GPD1-*aro1*$_{D1409A,D920A}$/TP1-*aro4*$_{K229L}$ /ADH1-*TKL1* / PGK1-*RKI1*
BY4741 **G1**	pRS415-MDH2	TEF1p-MDH2-HXT7t
	pRS413- SA_high	GPD1-*aro1*$_{D1409A,D920A}$/TP1-*aro4*$_{K229L}$ /ADH1-*TKL1* / PGK1-*RKI1*
BY4741 **G2**	pRS415-MDH2$_{P1S}$	TEF1p-MDH2(P1S)-HXT7t
	pRS413- SA_high	GPD1-*aro1*$_{D1409A,D920A}$/TP1-*aro4*$_{K229L}$ /ADH1-*TKL1* / PGK1-*RKI1*
BY4741Δ*oac1* **G3**	pRS415-MDH2	TEF1p-MDH2-HXT7t
	pRS413- SA_high	GPD1-*aro1*$_{D1409A,D920A}$/TP1-*aro4*$_{K229L}$ /ADH1-*TKL1* / PGK1-*RKI1*
BY4741Δ*oac1* **G4**	pRS415-MDH2 $_{P1S}$	TEF1p-MDH2(P1S)-HXT7t
	pRS413- SA_high	GPD1-*aro1*$_{D1409A,D920A}$/TP1-*aro4*$_{K229L}$ /ADH1-*TKL1* / PGK1-*RKI1*
BY4741::SA **G5**	pRS415-MDH2	TEF1p-*MDH2* $_{P1S}$ -HXT7t
BY4741::SA **G6**	pRS415-MDH2 $_{P1S}$	TEF1p-*MDH2* $_{P1S}$ -HXT7t

3.2. Fermentation conditions

Single yeast colonies were picked from the transformation plates and inoculated into 3 mL of SC-His-Leu media for analysis of SA production. After overnight growth, 30 µl of culture was transferred into 3 ml of SC-His-Leu (2% glucose) and after 72 h of fermentation, 1 ml samples were collected, centrifuged at 5000 rpm for 5 min, and the supernatant was stored at -20 ℃ until chromatographic analysis.

3.3. Biomass and metabolite analysis

Biomass growth in media was monitored by optical density changes with an absorbance reader at 600 nm (OD_{600}) (Biotek synergy 2 Multi-Mode microplate Reader). Analysis of SA was done by high liquid pressure liquid chromatography (HPLC) using an Agilent 1200 Series equipment (Agilent Technologies, USA) coupled to UV detector at 210 nm with an Aminex HPX-87H column (300×7.8 mm) (BioRad, Hercules, CA). The mobile phase was 0.5 mmol L^{-1} H_2SO_4 with the flow rate of 0.3 mL min^{-1}.

4. Results and discussion

Malate dehydrogenase (MDH2p) is the initial enzyme in the gluconeogenic pathway that catalyzes the reversible oxidation of malate to oxaloacetate in the cytosol. We hypothesized that overexpressing the MDH2 could enhance the pools of PEP leading to higher accumulation of SA. Unexpectedly, the production of SA slightly decreased when the MDH2p was overexpressed in the control BY4741 strain. A possible cause for this reduction could be that MDH2p favors the reverse reaction, which is the conversion of OAA to malate. In fact, previous reports show that increasing the concentration of this enzyme can lead to elevated titers of malate acid in *S. cerevisiae* (Zelle et al., 2008). To increase the availability and activity of MDH2p in the cytoplasm, we overexpressed a mutant version of this protein carrying a serine substitution in the proline 1 residue (Strain G2, **Table 1**). This mutation intended to remove the phosphorylation site in MDH2 which leads to protein degradation. As expected, this mutation reduced the degradation of MDH2 which led to increased activity and reduced SA production. The titers dropped 45% compared to the control strain. This indicated that

indeed, MDH2 favors the production of malate over OAA. We further studied the deletion of the OAA mitochondrial carrier *OAC1* to increase the levels of OAA in the cytoplasm. However, the production of SA decreased in this knockout strain, carrying either version of MDH2p, to less the 200 mg L^{-1} (**Figure 3**) with no correlation with a decrease in biomass production (**Figure 3**). Further studies should focus on overexpressing an NADPH malic enzyme to recover malate and convert it to pyruvate with the completion of the shunt by overexpressing a phosphoenolpyruvate synthase (*e.g. E. coli ppsA*). Another possibility could be overexpressing new enzymatic versions of MDH2p with known irreversible activity. One such example is the malate:quinone oxidoreductase from *E. coli* (*Ec_mqo,* **Figure 1**).

This rational can also be complemented by linking the metabolic engineering strategies with a biochemical driving force (Shen et al., 2011). Common driving forces are related energy exchange metabolites like ATP pools or other cofactors such as NADH and NADPH. In this light, the activity of the enzyme of study (present in the pathway of interest) can be elevated if its reaction can provide or complement a cofactor deficiency created by the deletion of selected genes. To enhance the production of aromatic compounds, such driving force can be created by deleting the *ZWF1* gene in the oxidative pentose phosphate pathway. This would require a new source for replenishment of cytosolic NADPH levels that can be attained by overexpressing dehydrogenase enzymes that depend on the $NADP^+$ cofactor. Overexpression of heterologous MDH2p enzymes that prefer $NADP^+$ as the cofactor could help establishing the metabolic driving force. One potential gene for this strategy is the NADPH-dependent MDH enzyme involved in pyruvate metabolism and carbon fixation pathways in several organisms such as *Zea mays* (Metzler et al., 1989), *A. thaliana* and *Methanobacterium thermoautotrophicum* (Thompson et al., 1998).

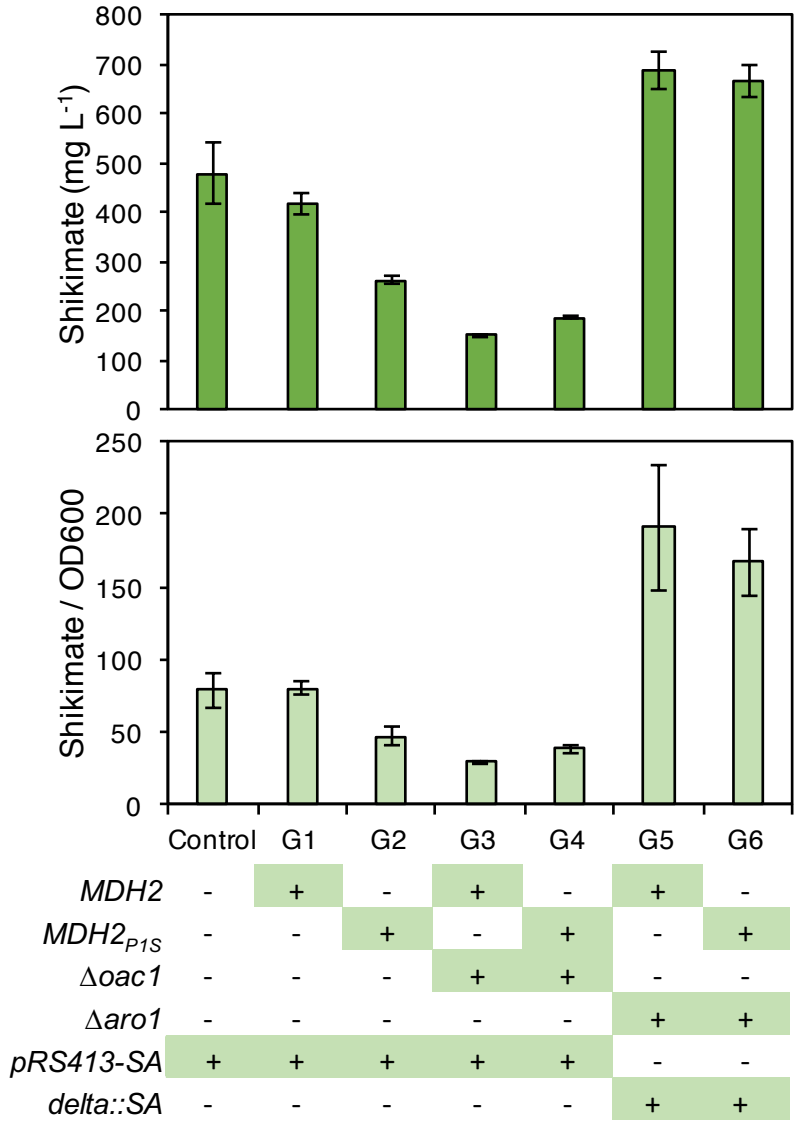


Figure 3. Fermentation results from strains engineered to produce SA. Samples were collected at 72 h. The variation indicates the standard deviation from three biological replicates.

5. New Products and New Species

5.1. Engineering the downstream module of the aromatic pathway in *Saccharomyces cerevisiae* and *Pichia stipitis*

As reviewed in Chapter 2, the aromatic amino acid pathway is a precursor to a vast diversity of high-value secondary metabolites. The efforts described in the subsequent chapters and in the previous section to increase the carbon entrance into the aromatic amino acid pathway have led to constructing high-producing strains for MA and SA accumulation. It is expected that the genetic platforms, as well as the metabolic engineering strategies demonstrated in previous sections, translate seamlessly for the downstream compounds. Moreover, the highest producers obtained thus far will serve as the starting point for the construction of new strains

The downstream module of the aromatic amino acid biosynthetic pathway encompasses two main branches: 1) the L-phe and L-tyr branch and 2) the L-trp branch. Most of the engineering work to produce downstream metabolites has focused on the L-tyr branch because it is the precursor to two big families of secondary metabolites, *i.e.*, flavonoids and stilbenoids, and benzylisoquinoline alkaloids. L-tyr can be converted to coumaric acid for production of nutraceuticals like naringenin and resveratrol. On the other hand, L-tyr can undergo deamination and hydroxylation to produce L-DOPA, which is the main precursor to the production of major pain-management drugs.

Although *S. cerevisiae* is the model yeast for unicellular eukaryotes, other yeasts possess diverse physiological features that make them more attractive for metabolic engineering applications. Such features include higher lipid accumulation such as *Yarrowia lipolytica* (Nicaud, 2012), higher assimilation rates of pentose sugars such as *Pichia stipitis* (Gao et al., 2016), or a Crab-negative physiology, which entails higher biomass formation in the presence

of elevated glucose, rather than the production of ethanol. In previous work, we have demonstrated that *P. stipits* is an excellent host for production of aromatic compounds due to its greater activity in the pentose phosphate pathway (Gao et al., 2016). Ultimately, the goal is to engineer this strain to exploit its potential in the production of valuable aromatics, specifically, those deriving from gene L-tyr.

5.2. Production of coumaric acid

The production of coumaric acid was established by overexpressing the gene encoding the tyrosine ammonia lyase (*TAL)* from *Flavobacterium johnsoniae* in an episomal vector. The product of this gene has been previously characterized and compared to other enzymes from the same class and has been demonstrated to possess the highest catalytic activity to convert tyrosine to coumaric acid (Jendresen et al., 2015).

A codon-optimized version of the *TAL* gene was synthesized for expression in *S. cerevisiae* and *Pichia stipitis*. The gene was cloned into the pRS414 vector under the expression of the strong promoter TEF1 and transformed into *S. cerevisiae.* For the production of coumaric acid in *P. stipitis*, the gene was cloned into an episomal vector for selection in tryptophan currently developed by our group. Transformation of the plasmids was done following the lithium acetate protocol and plated on SC-Ura plates. The selected colonies were grown in liquid selective media and supplemented with L-tyrosine to serve as the substrate for the tyrosine ammonia lyase enzyme. *S. cerevisiae* was able to produce 26.1 mg L^{-1} of coumaric acid while *P. stipitis* produced 13.3 mg L^{-1} (**Figure 4**). Both species produced coumaric acid at very similar yields, indicating that both enzymes have very similar activities. This is an indication that *P. stipitis* can be readily engineered to produce aromatic secondary metabolites.

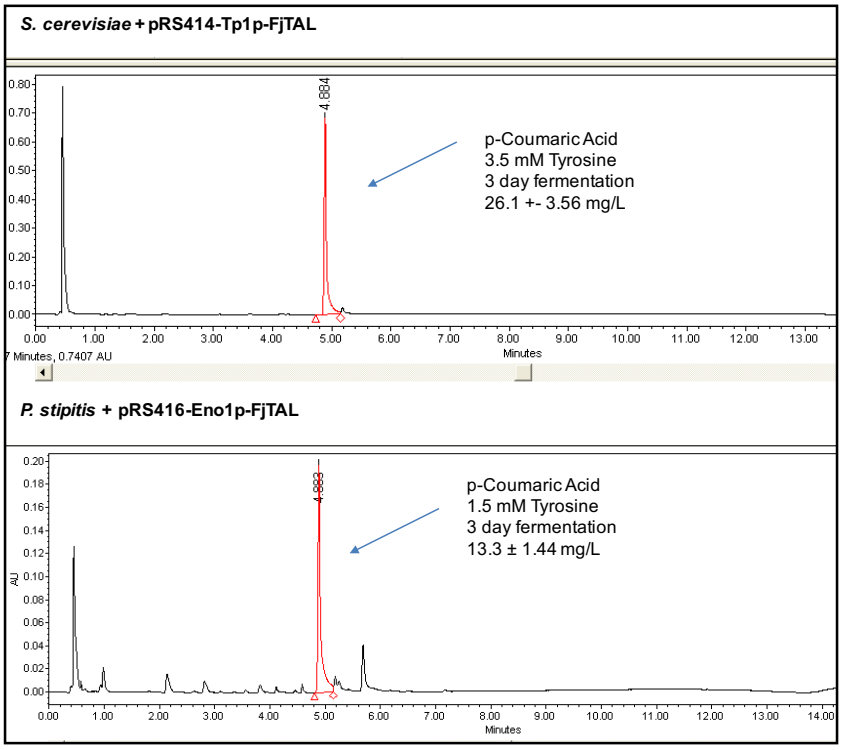


Figure 4: Chromatogram of a *S. cerevisiae* and *P. stipitis* engineered for the production of coumaric acid.

5.3. Engineering the downstream module of the aromatic amino acid pathway

The *S. cerevisiae* strain SA9 from the previous chapter, which was found to have the highest accumulation of shikimic acid (SA) will be used as platform strain for the production of coumaric acid. To further unlock the carbon flux into the downstream module, the *E. coli* gene *aroL* will be overexpressed. This gene codes for a shikimate kinase enzyme, which has been demonstrated to be a bottleneck step in *S. cerevisiae.* It is important to mention that the strain SA9 contains the Δ*aro1* which could be complemented with the *aroL* gene. However, we will also study the expression of this gene in a strain without the aforementioned deletion.

Feedback inhibition occurs also in the downstream aromatic module at the level of the chorismate mutase enzyme Aro7p. This enzyme undergoes inhibition by high levels of L-tyr

in order to keep a strict balance between the L-tyr/L-phe branch and the L-trp branch. Hence, we will overexpress a mutant version of this enzyme with the glycine residue 141 substituted to serine. This point mutation has been demonstrated previously to remove the feedback inhibition caused by L-tyr (Rodriguez et al., 2015). To avoid the loss of carbon through pathway competition, we will study the deletion of two genes, namely *ARO10* and *PDC5* which are known to initiate the degradation of the aromatic amino acids to use as the nitrogen source.

6. Future perspectives

6.1. Exploration of downstream transcriptional regulators

Following the studies described previously to find transcriptional repressors of the upstream module, new explorations can be performed to find new candidate factors that affect the downstream module of the aromatic amino acid pathway. This would entail analyzing the promoter sequences of the genes chorismate mutase (*ARO7*), prephenate dehydrogenase (*PHA2*), prephenate dehydrogenase (*TYR1*), and the two aromatic aminotransferases (*ARO8/9*).

This study can be approached in a combinatorial way and relying on a fast screening method to increase the overall efficiency. Initially, the knockout strains obtained from the YEASTRACT database can be pooled into a single dense yeast culture. Secondly, a minimal genetic platform composed of the structural genes and a gene for fast screening will be constructed in a plasmid containing the CRISPR system for gene integration, ensuring only a single transformation event per strain. After transformation, the pooled cells will be plated on selective media. The cells with a positive result according to the colorimetric biosensor will be

picked and back-sequenced to determine the locus of the gene knockout. This strategy is depicted in **Figure 5**.

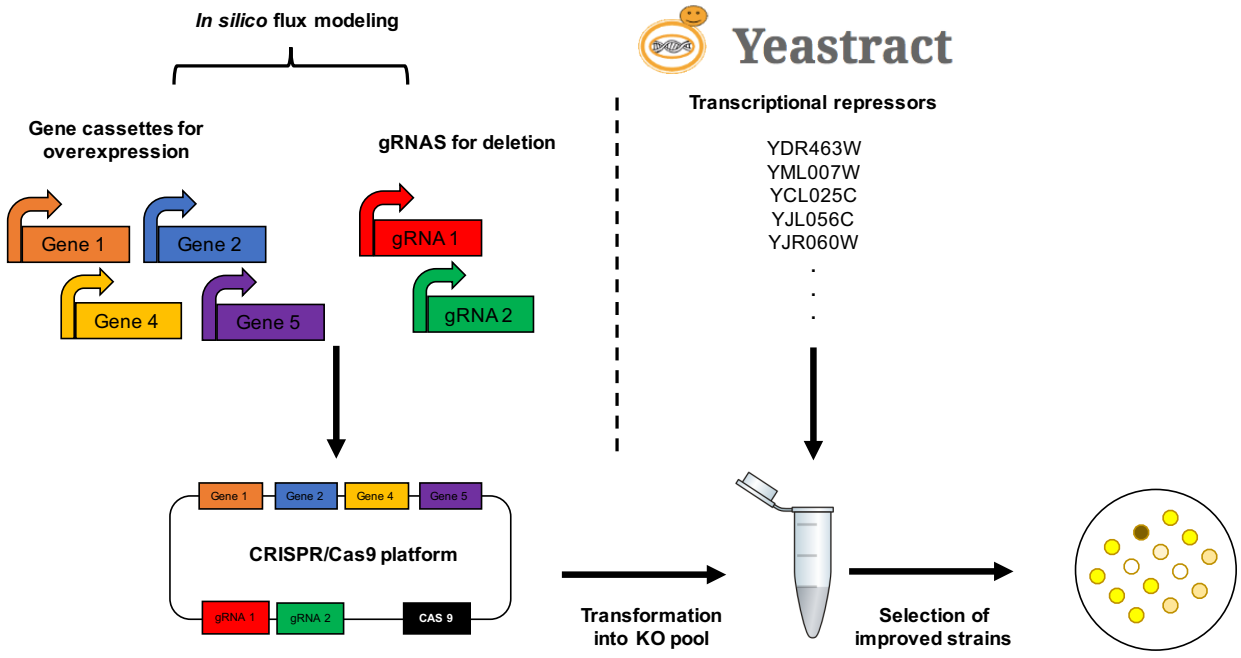


Figure 5: Depiction of proposed strategy for engineering the downstream aromatic amino acid branch.

6.2. CRISPR-dCas9-based metabolic valves to control the carbon flux into the aromatic pathway

In Chapter 5 OptForce analyses indicated a series of potential metabolic interventions that can be implemented to enhance the carbon flux into the aromatic amino acid pathway. Besides the overexpression of *RKI1*, the simulations suggested the deletion or downregulation of important genes in the glycolytic pathway (**Figure 3, Chapter 5**). The predicted gene targets included knocking out the phosphofructokinase gene (*PFK1*), and downregulating the genes phosphoglycerate kinase (*PGK*) and glylceraldehyde-3-phosphate dehydrogenase (*TDH*).

However, it is imperative to consider that the activity of these genes is crucial for utilization of fermentable sugars, such as glucose. Deleting these genes could result in very low biomass titers which would ultimately decrease the production of the desired compounds. Hence, new alternatives ought to be explored that aim to modulate gene expression rather than a strict knockout, and one such alternative could be achieved through the implementation of CRISPR-Cas technology.

The prokaryotic adaptive immune system, composed of Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes (CRISPR-Cas), is an immune system against bacteriophage infections. This widespread mechanism enables microbes to precisely recognize specific sequences of the invader DNA, and acquire immunity against future attacks (Makarova et al., 2011). From an evolutionary perspective, studying the structure and mechanisms of CRISPR-Cas systems has opened new avenues for understanding the interactions between prokaryotes and viral populations and how these have shaped many microbial ecosystems (Andersson and Banfield, 2008; Bondy-Denomy and Davidson, 2014). Furthermore, given the ability of the Cas proteins to cleave DNA with a high sequence specificity, CRISPR-Cas has been repurposed as a robust tool for gene editing. In fact, researchers have taken advantage of the fewer requirements of type II CRISPR systems and have exploited them for many elegant genome editing applications. In bacterial and eukaryotic cells, applying this powerful platform can be as simple as the insertion or deletion of genes. Other more sophisticated cases can be achieved by employing and engineered Cas9 lacking its nuclease activity (dCas9), but maintaining its DNA binding ability. This has opened the door for implementing genome-wide interference studies in several species (Hawkins et al., 2015; Kiani et al., 2015; Larson et al., 2013; Lv et al., 2015).

For instance, this technology can be implemented to control, in a combinatorial way, the transcription of more the one gene more metabolic engineering applications. To complement the predictions established by OptForce in the previous Chapter, such engineering platform would consist of a library of multi-loci gRNAs for precise binding of dCAS9 with modulated strength to the target genes of interest. The second component is an optimized dCAS9 engineered for different applications, including but not limited to: (i) gene silencing by promoter blockage, (ii) gene activation when linked to a transcriptional activator, or (iii) gene modification when linked to a transferase-like catalyst. This platform would enable the modulation of gene expression within its native genome context, bypassing the need to utilize heterologous or synthetic promoters. Rather than limiting the system to an "on" or "off" state (*i.e.,* transcription vs. no transcription), it can provide a platform in which genes can be studied in a wide range of transcriptional strengths. By linking the gene of interest to a reporter, the screening process can be simplified, although this is not a requirement, *e.g.*, when the gene of interest provides an evident visual phenotype, or when a high throughput screening method is in hand.

This CRISPR-based tool can be exploited to study phenotypes resulting from manipulating the transcriptional strength of more than one gene simultaneously, for example, to downregulate the genes in glycolysis to enhance the production of aromatic compounds (*TDH*, *PGK1*, *PFK1*). This property enables the creation of genetic landscapes that otherwise would not be possible by single gene deletion or overexpression. The "fluidity" property of the transcriptional strength, or partial loss/gain of function, provided by CAMEO-dCas9 is of special interest when the deletion of the genes of study is lethal. This platform allows for a fine-tuning of gene expression. **Figure 6** describes the construction and application of this

platform to target more than one gene simultaneously. The construction of the library of multiplex gRNAs can be performed by rapid *in vivo* assembly methods such as DNA assembler (Shao et al., 2009).

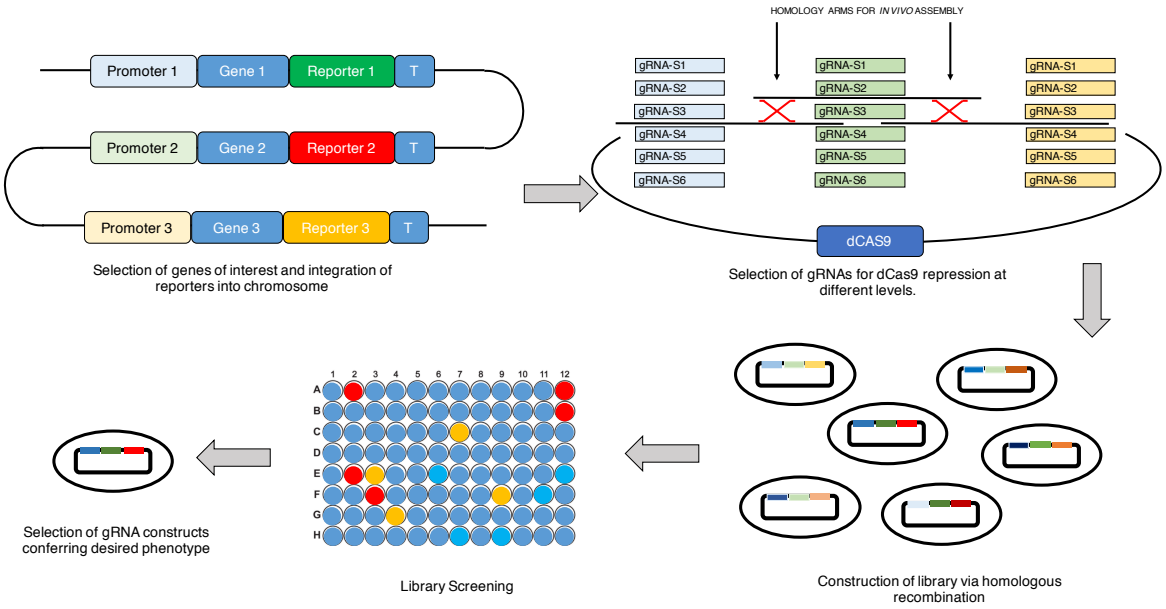


Figure 6: Schematic representation of a CRISPR-Cas9 based platform for dynamic regulation of genes. The genes of interest are linked to a reporter gene for ease of screening; this can be bypassed when high-throughput methods are available for the desired phenotype. The combinatorial, *in vivo*-constructed gRNA library, is then introduced into the host cell. Selection of desired cell lines or strains is followed by sequencing of the gRNA construct that confers the phenotype of interest. The system is not limited to the number of genes of interest or type of dCas9; gene activation (Maeder et al., 2013*)* and modification, rather than repression, can be achieved by fusing dCas9 to other catalytically active units.

7. Conclusions

Engineering the aromatic amino acid pathway in *S. cerevisiae* has proven to be a challenging endeavor. Nevertheless, successes in the production of important molecules like shikimic acid and muconic acid has been achieved through the implementation of several *state of the art* tools and technologies. As a matter of fact, by developing innovative engineering strategies, this work has built upon previous literature to construct strains with the highest reported yields and titers in molecules deriving from the pathway of interest. This work contains a compilation of studies that contribute towards understanding the metabolism of aromatic compounds in *S. cerevisiae* and establishes new engineering strategies that can be further implemented for the construction of new strains to manipulate new pathways.

To establish successful metabolic engineering strategies to exploit the production of aromatic derivatives, it was imperative to comprehend the two main core metabolic pathways that provide the precursor substrates. The pentose phosphate pathway (PPP) serves an important function in the cell as it constitutes the main source of reducing power in the form of NADPH units. The gene *ZWF1*, a major component of the oxidative portion of PPP, catalyzes the dehydrogenation of glucose-6-phosphate utilizing $NADP^+$ as a cofactor. Although several reports indicate that deleting this gene could enforce the carbon entrance through the non-oxidative PPP to allow a direct increase in E4P pools, major rearrangements occur in the cell to find new NADPH sources. This lowers the cells' fitness and may not be an effective strategy for the establishment of industrial microbial factories.

More dynamic approaches need to be developed to allow engineering the partition of the flux between two of the most important pathways in *S. cerevisiae*, namely, glycolysis and the

pentose phosphate pathway for the accumulation of aromatic amino acid compounds. With new advancements in gene editing technology, it is not unlikely that new strategies to explore large phenotypic landscapes will be constructed to rationally engineer the metabolism for the production any desired compound. Current *in silico* methods to predict gene deletion and overexpression may be greatly complemented by synthetic biology approaches that allow fast validations of the computational predictions. However, it remains unknown whether *S. cerevisiae* is the best microorganism to allow the industrial production of aromatic amino acid derivatives. As presented in this chapter and in previous publications, the unconventional yeast *P. stipitis* is a great candidate for the biosynthesis of compounds from this family, however, the tools available to engineer this yeast need to be further optimized to allow the establishment of high-producing strains. In the meantime, *S. cerevisiae* still remains as an organism that can be used as an exploratory model for metabolic engineering studies. It is expected that the engineering rationale should be translatable among close-related species, however, intrinsic differences such as regulatory networks at all levels of metabolism may remain as an additional challenge to overcome.

The following list of statements summarize the conclusions obtained from this work:

- The production of the target metabolite may not be fully accomplished through a biological route. In this case, a chemical step for diversification may be required. It is recommended to simultaneously evaluate the parameters of the main catalytic steps to enable seamless process integrations and to avoid expensive separation steps.
- Although *S. cerevisiae* is the most studied yeast, several unknown factors can result in drastic production differences when engineering different wildtype strains from

the same species. Such strain-dependent phenomenon can be taken advantage of to develop new metabolic engineering rationales.

- Multilevel metabolic engineering strategies following an orderly algorithm should allow high production titers of the target compound.
- Understanding the transcriptional regulation that affects specific pathways is determinant in unlocking higher production titers. Promoter sequence analysis of closely related genes can represent a starting point to obtain important insights into the regulation of a specific pathway.
- Single-gene perturbations may not be enough to understand the metabolic landscape within the cell, hence new approaches need to be devised targeting global perturbations.
- *In silico* models that predict gene perturbations and carbon flux distributions require fast synthetic biology strategies for validation and high-throughput detection methods.
- Regulations at all metabolic levels (transcriptional, post-transcriptional, post-translational, and allosteric) should be expected and explored to engineer central and energy-demanding pathways.
- The construction of microbial factories must not be limited to common model microorganisms. Non-conventional yeasts, with exemplary physiological features, may represent more feasible alternatives to achieve higher yields and titers.

8. Bibliography

Albuquerque, C. P., Smolka, M. B., Payne, S. H., Bafna, V., Eng, J., Zhou, H., 2008. A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics. 7**,** 1389-96.

Amberg, D. C., Burke, D. J., Strathern, J. N., 2006. "Quick and dirty" plasmid transformation of yeast colonies. CSH Protoc. 2006

Andersson, A. F., Banfield, J. F., 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science. 320**,** 1047-50.

Bondy-Denomy, J., Davidson, A. R., 2014. To acquire or resist: the complex biological effects of CRISPR-Cas systems. Trends Microbiol. 22**,** 218-25.

Cardenas, J., Da Silva, N. A., 2016. Engineering cofactor and transport mechanisms in *Saccharomyces cerevisiae* for enhanced acetyl-CoA and polyketide biosynthesis. Metab Eng. 36**,** 80-9.

Gao, M., Cao, M., Suastegui, M., Walker, J. A., Rodriguez-Quiroz, N., Wu, Y., Tribby, D., Okerlund, A., Stanley, L. M., Shanks, J. V., Shao, Z., 2016. Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. ACS Synth Biol. 6**,** 29-38.

Hawkins, J. S., Wong, S., Peters, J. M., Almeida, R., Qi, L. S., 2015. Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi). Methods Mol Biol. 1311**,** 349-62.

Hung, G. C., Brown, C. R., Wolfe, A. B., Liu, J., Chiang, H. L., 2004. Degradation of the gluconeogenic enzymes fructose-1,6-bisphosphatase and malate dehydrogenase is mediated by distinct proteolytic pathways and signaling events. J. Biol. Chem. 279**,** 49138-50.

Jendresen, C. B., Stahlhut, S. G., Li, M., Gaspar, P., Siedler, S., Forster, J., Maury, J., Borodina, I., Nielsen, A. T., 2015. Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and *Saccharomyces cerevisiae*. Appl Environ Microbiol. 81**,** 4458-76.

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S., Buchthal, J., Kowal, E. J., Ebrahimkhani, M. R., Collins, J. J., Weiss, R., Church, G., 2015. Cas9 gRNA engineering for genome editing, activation and repression. Nat Methods. 12**,** 1051-4.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., Qi, L. S., 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. 8**,** 2180-96.

Lv, L., Ren, Y. L., Chen, J. C., Wu, Q., Chen, G. Q., 2015. Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metab Eng. 29**,** 160-8.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., Joung, J. K., 2013. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 10**,** 977-9.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J., Koonin, E. V., 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 9**,** 467-77.

Metzler, M. C., Rothermel, B. A., Nelson, T., 1989. Maize NADP-malate dehydrogenase: cDNA cloning, sequence, and mRNA characterization. Plant Mol Biol. 12**,** 713-22.

Minard, K. I., McAlister-Henn, L., 1992. Glucose-induced degradation of the MDH2 isozyme of malate dehydrogenase in yeast. J. Biol. Chem. 267**,** 17458-64.

Nicaud, J. M., 2012. *Yarrowia lipolytica*. Yeast. 29**,** 409-18.

Palmieri, L., Vozza, A., Agrimi, G., De Marco, V., Runswick, M. J., Palmieri, F., Walker, J. E., 1999. Identification of the yeast mitochondrial transporter for oxaloacetate and sulfate. J. Biol. Chem. 274**,** 22184-90.

Patnaik, R., Liao, J. C., 1994. Engineering of Escherichia coli central metabolism for aromatic metabolite production with near theoretical yield. Appl Environ Microbiol 60**,** 3903-8.

Reider Apel, A., d'Espaux, L., Wehrs, M., Sachs, D., Li, R. A., Tong, G. J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N. J., Keasling, J. D., Mukhopadhyay, A., 2017. A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. Nucleic Acids Res. 45**,** 496-508.

Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., Nielsen, J., 2015. Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. Metab Eng. 31**,** 181-188.

Schuller, H. J., 2003. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. Curr Genet. 43**,** 139-60.

Shao, Z., Zhao, H., Zhao, H., 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37**,** e16.

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., Liao, J. C., 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. Appl. Environ. Microbiol. 77**,** 2905-15.

Suástegui, M., Guo, W., Feng, X., Shao, Z., 2016a. Investigating strain dependency in the production of aromatic compounds in *Saccharomyces cerevisiae*. Biotechnol Bioeng. 113.

Suástegui, M., Matthiesen, J. E., Carraher, J. M., Hernandez, N., Rodriguez Quiroz, N., Okerlund, A., Cochran, E. W., Shao, Z., Tessonnier, J. P., 2016b. Combining metabolic engineering and electrocatalysis: application to the production of polyamides from sugar. Angew. Chem., Int. Ed. 55**,** 2368-73.

Suástegui, M., Shao, Z., 2016. Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. J Ind Microbiol Biotechnol. 43**,** 1611-1624.

Swaney, D. L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N. J., Villen, J., 2013. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat Methods. 10**,** 676-82.

Thompson, H., Tersteegen, A., Thauer, R. K., Hedderich, R., 1998. Two malate dehydrogenases in *Methanobacterium thermoautotrophicum*. Arch Microbiol. 170**,** 38-42.

Yin, Z., Hatton, L., Brown, A. J., 2000. Differential post-transcriptional regulation of yeast mRNAs in response to high and low glucose concentrations. Mol Microbiol. 35**,** 553-65.

Zelle, R. M., de Hulster, E., van Winden, W. A., de Waard, P., Dijkema, C., Winkler, A. A., Geertman, J. M., van Dijken, J. P., Pronk, J. T., van Maris, A. J., 2008. Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. Appl Environ Microbiol 74**,** 2766-77.

APPENDIX I

COMMON MOLECULAR BIOLOGY TECHNIQUES FOR PATWHAY CONSTRUCTION AND OPTIMIZATION

A modified version of the book chapter published in the 1st Edition of Biotechnologies for Biofuel Production and Optimization, Elsevier.

Authors: Miguel Suástegui[1], Meirong Gao[1], and Zengyi Shao[1]
[1]Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

1. Introduction

In this thesis, the metabolic engineering were largely based on *in silico* modeling of carbon fluxes and strongly complemented with efficient synthetic biology techniques for fast assembly of DNA platforms. Assembly methods such as Gibson Assembly, Golden Gate Assembly, and DNA Assembler were implemented throughout this work. Hence, this appendix is dedicated to summarizing these outstanding technologies and expands on their applications to engineer microbial factories. Overall, this appendix may serve as reference for the establishment of assembly and optimization of genetic platforms for engineering microorganisms to produce compounds that go beyond the aromatic amino acid pathway.

The technology described in this appendix encompasses trending methods for pathway engineering and optimization. Although efficient metabolite production in microbes does not rely solely on the introduction of heterologous pathways, it represents a big part of the engineering foundation to obtain industrial-level microbial factories. Nonetheless, to increase the pace of these advancements, several key aspects have to be enabled. One of them is to develop truly combinatorial assembly methods, *i.e.*, methods that allow mixing and matching genetic parts in a few steps while relying on inexpensive chemicals. Developing fast screening techniques is also indispensable for the efficient selection of mutants with desired phenotypes.

And finally, it is important to develop assembly techniques capable of being adopted by a programing language for implementation in liquid-handlers or microfluidic devices. By transferring as most wet-lab work to a machine as possible, we are reducing potential errors, time, and money expenditures. But most importantly, we are ensuring that other researchers can easily adopt such technologies, which opens the door for fast advancements in the already palpable aims to substitute petroleum-derived chemicals with its biological counterparts.

Pathway assembly represents one of the key aspects in the engineering of microbes for the production of valuable chemicals. With the aid of synthetic biology, recombinant DNA technology, and bioinformatics (Kahl and Endy, 2013; Priscilla and Ron, 2009), nowadays it is possible to assemble finely tuned pathways capable of producing a wide variety of natural and non-natural products (Lee et al., 2008; Rabinovitch-Deere et al., 2013) . The concept of pathway assembly has exponentially evolved with the development of tools that range from efficient and cheap DNA synthesis (Kahl and Endy, 2013; Robert, 2009) to microfluidics. At the present time, what used to be a time-consuming endeavor is now limited by one's imagination.

There is a commonly accepted "algorithm" for designing heterologous biosynthetic pathways (**Figure 1**). When a target molecule is identified, the first step is to check whether there are natural pathways responsible for its biosynthesis. If the enzymes have not been previously characterized, homolog search (Boratyn Grzegorz et al., 2012) is usually performed to help discern the enzymes that could best carry the desired reactions. In this step, phylogenesis (Jing et al., 2011) of the source organisms and the desired production host is taken into consideration. The functional expression of "difficult enzymes" is more likely to happen if two hosts have a closer evolutional relationship. With this, a group of candidate

enzymes can be selected to perform the desired set of reactions towards the molecule of interest. In case that a target molecule does not have precedence in nature, a biosynthetic pathway needs to be designed *de novo*. Generally, this process starts with searching for a pathway that synthesizes a product or an intermediate similar to the target compound in structure (Bloch and Schmidt-Dannert, 2014; Clementina et al., 2011; Ren et al., 2010). Next, how to design feasible routes to complete the remaining conversions is usually challenging and mostly important. Enzyme promiscuity sometimes allows for a direct fusion of two independent pathways; if not, protein engineering needs to be applied, rendering novel activities to the existing biocatalysts to complete the desired conversions (Lane and Seelig, 2014; Li and Cirino, 2014; Pandya et al., 2014).

The next important step is to select the organism in which the pathway will be implemented. This represents a pivotal point in the design of the pathway. Considerations such as codon-optimization, transcriptional and translational regulation will have to be taken into consideration. The battle horses for heterologous pathway expression are usually the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae,* due to their well-understood metabolism and physiology when compared to other organisms. Great efforts have been put into understanding the mechanisms of gene expression and regulation in these organisms, allowing the development of powerful tools that permit quick manipulations in these hosts (Redden et al., 2014). Nevertheless, issues like low tolerance to exogenous metabolites have become a drawback in the implementation of metabolic pathways in such model microorganisms. This is mostly common in the heterologous production of biofuels and chemicals (Royce et al., 2013). Developing strains with higher tolerance to these metabolites are a strong area of research. The advent of technologies like genome sequencing and genome

editing have allowed the manipulation of non-conventional microorganisms with outstanding features that make them perfect candidates for the production of toxic compounds (Xiao et al., 2014).

Once the host organism has been selected, computational program such as Vector NTI (Lu and Moriyama, 2004), j5 (Hillson et al., 2012), and Gene Designer (Villalobos et al., 2006) can be implemented for convenient visualization of DNA parts. The next step consists of assemble the DNA pieces either *in vitro* or *in vivo*. Numerous techniques have been developed to carry out this task, all with the same objective: to maximize the assembly efficiency. The assembly efficiency will consist in, among other things, the time required to finalize pathway assembly with minimum errors.

To finalize the pipeline, the pathway is introduced into the host organism and tested for the desired metabolite using, most commonly, chromatographic tools. After the initial characterization, the pathway may be modified in order to alter, for example, regulatory elements or the initially selected structural genes. Combinatorial design has become a new trend in pathway optimization, allowing the generation of large libraries of pathways with varying features. Elegant techniques have emerged from the need of generating strains with elevated titers in the case of biofuels; however, these techniques can also be applied to engineer the production of non-natural compounds (Cobb et al., 2014; Du et al., 2012).

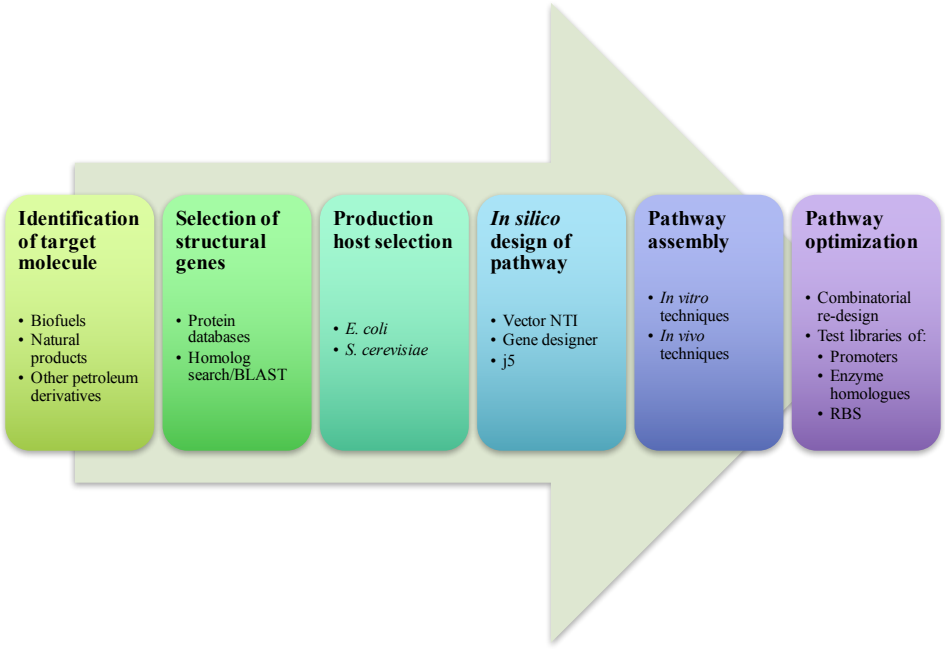


Figure 1: Common rationale for pathway design.

2. Assembly methods

This section describes the most commonly used DNA assembly methods. **Table 1** summarizes these techniques depicting the key features and the largest construct that was obtained as reported in the corresponding reference.

2.1. *In vitro* assembly based on homologous recombination

Unlike the conventional restriction enzyme- and ligase-dependent DNA cloning method, assembly based on homologous recombination does not require specific restriction sites. This enables the assembly of multiple fragments in a single step. By harnessing the power of homologous recombination, constructs can be assembled *in vitro* from a mixture of DNA fragments with overlapping regions. The assembled plasmids are then introduced into bacteria for enrichment, which finalizes the regeneration of the construct to double-stranded, intact

molecules. Among this category, SLIC, Gibson isothermal assembly, UNS, and PaperClip are typical and commonly used approaches

2.1.1. SLIC

Li *et al.* developed a sequence- and ligation-independent cloning (SLIC) method in 2007 (Li and Elledge, 2007). SLIC employs the exonuclease activity of T4 DNA polymerase to chew back the ends of an insert and a vector to reveal 5' single-stranded overhangs. Two single-stranded stretches will subsequently anneal based on homology to form a circular molecule. Addition of bacterial recombinase, RecA, could facilitate the annealing efficiency. Following electroporation to *E. coli*, the gaps in the recombinant molecule can be repaired by the endogenous homology recombination enzymes such as resolvases, nucleases, and polymerases. To stimulate an efficient recombination *in vitro*, 30-40 bp homology overlapping regions and a molar ratio of 2-4:1 between an insert and a vector are recommended.

To prove the efficiency of SLIC, as many as ten fragments with 40-bp overlapping regions were assembled to generate a circular plasmid carrying a kanamycin-resistant gene. Among the fragments, nine were PCR-generated DNA molecules with sizes ranging from 275 bp to 980 bp and the other one was a 3.1-kb linearized plasmid. It was observed that almost 20% of the colonies obtained after transformation had correct restriction digestion pattern (Li and Elledge, 2007).

.

Table 1: Summary of pathway assembly methods.

Methods	Main characteristics	Size of the largest pathway reported (kb)*	Uniqueness	Reference
SLIC	*In vitro* homology-based assembly with sticky ends	10.2		(Li and Elledge, 2007)
Gibson Isothermal Assembly		583	Commercial kit available	(Gibson et al., 2009)
BioBrick/BglBrick		~7	Commercial kit available	(Anderson et al., 2010)
UNS-guided Assembly		64	Algorithm available; Can be used for assembly of sequences with high similarity	(Guye et al., 2013)
SLiCE		24	Facilitated by bacterial cell extracts	(Zhang et al., 2012)
USER		3.2		(Bitinaite et al., 2007)
PIPE		7.0		(Stevenson et al., 2013)
LCR		20		(de Kok et al., 2014)
CPEC		8.4		(Quan and Tian, 2009)
PaperClip	*In vitro* homology recombination-based assembly	6.0		(Trubitsyna et al., 2014)
TAR	*In vivo* homology recombination-based assembly	250	Use *S. cerevisiae* as an assembly platform	(Natalay and Vladimir, 2006)
DNA Assembler		52		(Shao et al., 2013)
Genome Assembly		(1.08 Mb genome)		(Annaluru et al., 2014)
MAGIC		3.9	Assembly accomplished through bacterial mating	(Li and Elledge, 2005)
Golden Gate	*In vitro* Type II restriction enzyme digestion; Particularly useful for assembly of highly repeated fragments	33		(Weber et al., 2011a)
MASTER		36.4		(Chen et al., 2013)
GoldenBraid Assembly		14.3		(Sarrion-Perdigones et al., 2011)
Pairwise selection assembly		91		(Blake et al., 2010)
OGAB		13	Use *B. subtilis* as an assembly platform	(Tsuge et al., 2003)

2.1.2. Gibson Isothermal Assembly

Gibson isothermal assembly is one of the most widely used DNA assembly methods due to its ease-of-use, flexibility, and capacity for constructing large vectors (**Figure 2**). It simplifies the recombination of overlapping DNA molecules in a single step while maintaining high efficiency. Very similar to SLIC, the assembly process begins with an enzyme cocktail (including an exonuclease, a DNA polymerase, and a DNA ligase) mixed with a set of overlapping DNA fragments. The function of the exonuclease is to remove nucleotides from the 5' ends of double-stranded DNA molecules, thus exposing complementary single-stranded 3' overhangs for homologous annealing. The DNA polymerase fills in the remaining gaps after annealing, and the DNA ligase seals the nicks in the assembled DNA construct. After a short incubation period, double-stranded fully sealed DNA molecules are obtained. Gibson Assembly® Master Mix is commercially available, making this technique amenable to high-throughput assembly automation.

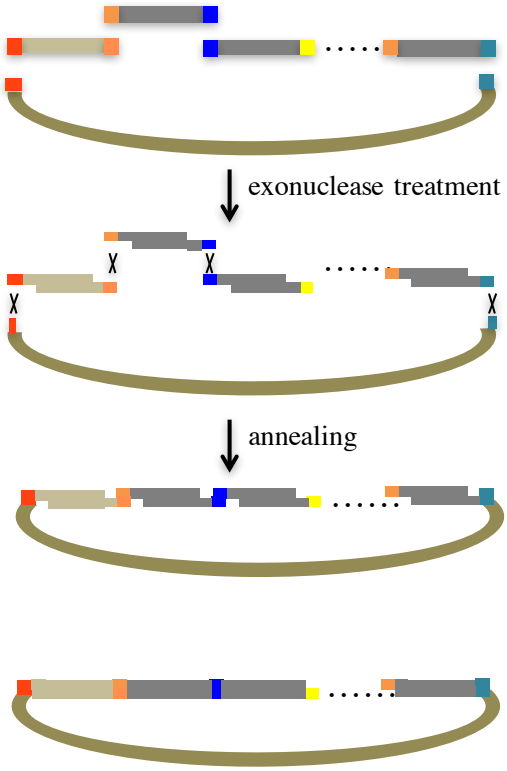


Figure 2. Gibson Isothermal Assembly *in vitro*: Exonuclease is employed to generate ssDNA overhangs. Polymerase and ligase are used to repair annealed fragments and fill gaps

2.1.3. UNS

On the basis of Gibson Assembly, Unique Nucleotide Sequence (UNS) assembly introduces a computational algorithm to design short unique overlaps (40 bp) between adjacent DNA fragments. This allows to drive an ordered assembly with high efficiency, especially when repeated sequences are involved (Torella et al., 2014). The algorithm is built upon the following premises: UNSs are balanced on GC content, have no secondary hairpin structure, avoid promiscuous hybridization, and evade homology to start codons, promoter-like sequences, and commonly used restriction sites. Although the principle guiding UNS assembly is identical to Gibson Assembly and very similar to SLIC, it uniquely offers standard overlapping parts that can be easily repurposed among synthetic pathways and circuits. Therefore, this technology promotes a philosophy of standardization and insulation as emphasized in synthetic biology. To append the UNSs to the DNA fragments to be assembled, they can be cloned from preassembled vectors, from PCR amplification with primers containing the UNSs, or simply by direct synthesis of the cassettes with the corresponding UNSs.

2.1.4. PaperClip

PaperClip is a DNA assembly method based on bridging oligos attached to the ends of the fragments to be assembled. This flexible technique allows the assembly of DNA fragments without the need of restriction digestion, or specific primers to add end-homology. The attached oligos (clips) drive an orderly assembly process, enabling combinatorial assembly of fragments (Trubitsyna et al., 2014).

Four single stranded oligos are required for every part to be assembled. These oligos contain homology with the assembled part as well as 3-bp sticky ends to allow the formation of a full clip. Pairs of upstream oligos and downstream oligos of each DNA part are first annealed separately to form the upstream and downstream half-Clip. Then the downstream half-Clip of one DNA molecule and the upstream half-Clip of its adjacent downstream molecule, sharing the designed complementary tri-nucleotide sequence, are ligated to form an intact clip. The assembly reaction containing a destination backbone, DNA fragments, and clips can be performed either in cell extract of *E. coli* expressing recombination protein or through two-step PCR to maintain high efficiency.

The key highlight of this method is the direct assembly of existing DNA parts without further cloning and PCR amplification with the help of four single stranded oligonucleotides.

Moreover, the four single stranded oligonucleotides, the half-clips, and the full clips can be stored and reused to the user's convenience. This feature is a great advantage in automated DNA assembly processes.

2.2. *In vivo* assembly based on homologous recombination

A series of assembly methods based on *in vivo* homologous recombination have been developed. These include the transformation associated recombination (TAR) (Ma et al., 1987), DNA Assembler (Shao et al., 2009), and mating-assisted genetically integrated cloning (MAGIC) (Li and Elledge, 2005), all of which take advantage of the endogenous homologous recombination machinery in biological cells (**Table 1**).

2.2.1. DNA Assembler

Shao *et al.* developed a powerful assembly tool, named DNA assembler, which allows single-step assembly of biochemical pathways *in vivo* (Shao et al., 2012; Shao et al., 2009). This method relies on the native homologous recombination machinery of *S. cerevisiae* to anneal multiple overlapping DNA sequences. The length of overlaps should be at least 40 bp to maintain high efficiency. In this approach, multiple target DNA fragments and a linearized vector are co-transformed into *S. cerevisiae* through electroporation (**Figure 3a**). The transformants are selected from amino acid dropout media according to the auxotrophic marker carried by the vector, and the correct constructs are confirmed by restriction digestion or sequencing. This method usually takes one more week than *in vitro* assembly because the assembled plasmids in yeast need to be retransformed to *E. coli* for DNA enrichment before being confirmed. However, because the assembly is conducted inside yeast cells, the power of *S. cerevisiae* homology recombination can be fully picked up, making the assembly accuracy and capacity significantly higher than any of the *in vitro* assembly methods. It has been demonstrated to construct a plasmid as large as 52 kb (Shao et al., 2013) although the upper limit could be far beyond as indicated in genome assembly based on the same principle. Shao *et al*. also described the versatility of this approach in manipulating large natural product gene clusters, including achieving heterologous expression, single or multiple point mutations, scar-less gene deletion, synthesis of natural product derivatives (Shao et al., 2011; Shao et al., 2012) and artificial gene clusters refactoring (Luo et al., 2013; Shao et al., 2013).

In addition, DNA assembler also enables direct chromosome-based assembly if the two outermost fragments are designed to anneal to the selected locus on the yeast chromosome

(**Figure 3b**). More of this technique in the context of combinatorial pathway assembly is described in the Section 4 of this appendix.

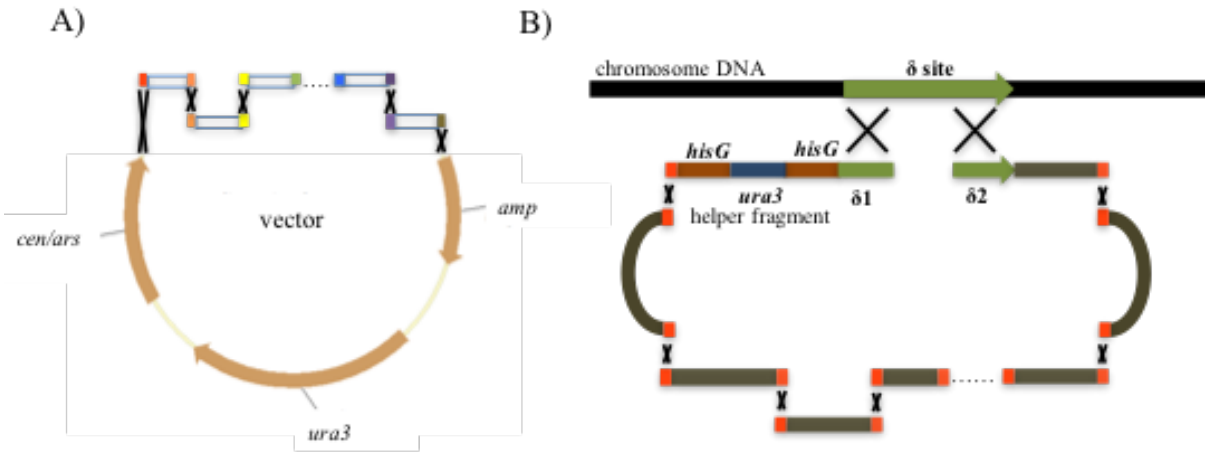


Figure 3. DNA Assembler based on *in vivo* assembly: A) Vector-based assembly; B) Chromosome-based assembly.

2.2.2. Genome assembly

A key milestone in DNA assembly history was achieved in 2008 with the invention of the Gibson Assembly. Daniel Gibson reconstituted the complete synthetic 583-kb genome of *Mycoplasma genitalium* (Gibson et al., 2008). The whole genome was split into 101 fragments with sizes ranging from 5 kb to 7 kb and overlapping regions ranging from 80 bp to 360 bp. Four quarter-genomes were individually assembled through Gibson Assembly, and then transferred to *S. cerevisiae* to reconstitute the whole genome *in vivo*. It is the yeast *in vivo* homologous recombination that ensures such a superior efficiency for large DNA fragment assembly when compared to *in vitro* assembly.

In 2010, Gibson *et al*. further assembled the 1.08-Mb *Mycoplasma mycoides* genome and more importantly, transplanted it to the "shell" of *Mycoplasma capricolum* whose genome had

been removed(Gibson et al., 2010). The newly transplanted cell demonstrated the phenotype directed by the synthesized genome rather than the recipient cell and was capable of continuous self-replication. Beyond the numbers game, the significance of this achievement is, for the first time, life was synthesized purely based on chemical building blocks. It can be imagined that, in the future, valuable creatures can be digitally designed and created according to human desires, despite numerous bioethical controversies it will likely initiate.

The most recent groundbreaking progress in genome assembly has been made through synthesizing the first eukaryotic chromosome*, S. cerevisiae* chromosome III, and moreover removing dispensable junk sequences to give a remarkable 14% reduction in chromosome size (Annaluru et al., 2014). Starting off with the efforts of 60 undergraduate students enrolled in the "Build a Genome" project seven years ago, the project has quickly sparked a global collaboration, named Sc2.0. This initial success in minimizing one yeast chromosome without losing major functions will lead to the reconstruction of the other 15 yeast chromosomes. Once completed, Sc2.0 will push the boundaries of synthetic biology through creating synthetic yeast with a minimized genome that will bolster development of more efficient industry processes for producing biofuels, chemicals and vaccines.

2.2.3. Non homology-based assembly

Non homology-based assembly relies on type II restriction enzymes or serine integrases that recognize and cleave specific DNA molecules at fixed sites. This leads to 5' or 3' DNA overhangs with potentially infinite combination of nucleotides. On the basis of this property, a series of pathway assembly tools have been built such as Golden Gate Assembly(Engler et al., 2009; Engler et al., 2008), Ordered Gene Assembly in *Bacillus subtilis* (OGAB) (Tsuge et

al., 2003), Serine integrase recombinational assembly (SIRA) (Colloms et al., 2014), GoldenBraid (Engler et al., 2008), and the pairwise selection assembly method (**Table 1**).

2.2.4. Golden Gate Assembly

The uniqueness of Golden Gate Assembly is the recognition sites of type II restriction enzymes which are removed during digestion, therefore allowing digestion and ligation to take place in the same pool (Engler et al., 2008). As shown in **Figure 4**, the fragments to be assembled are flanked with type II cleavage sites and the recognition sites. Incubation of these fragments with DNA ligase and a single type II restriction enzyme leads to an irreversible removal of the recognition sites. By designing unique overhangs, Golden Gate Assembly allows an orderly assembly of multiple fragments simultaneously. It has been shown that nine fragments were efficiently assembled into the recipient vector using Golden Gate method (Engler et al., 2009). More importantly, unlike homology-based assembly, Golden Gate Assembly has been proven to be particularly useful for building large genes or pathways with repeated sequences. For example, it was used to prepare transcription activator-like effector nucleases (TALENs) for genome editing, whereas homology-based assembly methods were useless because TALENs usually have 12-23 repeat modules (Weber et al., 2011b).

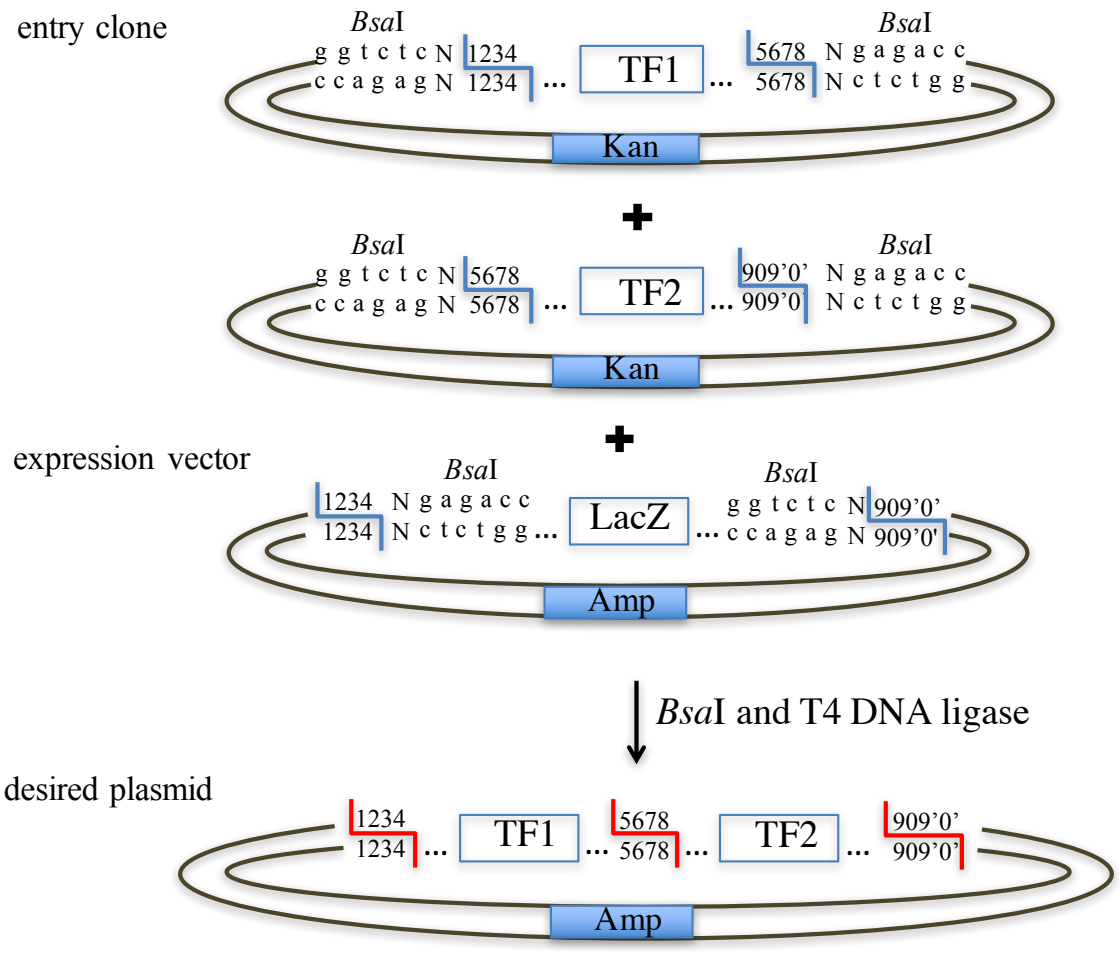


Figure 4: Golden Gate Assembly: The final desired plasmid is resistant to *Bsa*I digestion. Blue elbow connectors show the *Bsa*I restriction digestion sites. TF, target fragment. Number 0 to 9 (including 0' and 9') denote any nucleotide sequence.

2.2.5. OGAB

Besides *E. coli* and *S. cerevisiae*, other microorganisms possessing unique characters beneficial to DNA assembly have been subject of intense research (Cao et al., 2014; Gao et al., 2014; Tsuge et al., 2003). Ordered Gene Assembly in *Bacillus subtilis* (OGAB) is one of these examples(Tsuge et al., 2003). Unlike *E. coli*, *B. subtilis* prefers to uptake linear-form DNA and has nucleases located on its cell membranes to process the attached double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA). It can incorporate as many as 50 ssDNA per cell simultaneously. Once inside the cell, the single DNA strands will pair with their complementary fragments to form desired circular plasmids. This is highly facilitated by the endogenous recombination machinery in *B. subtilis*. In fact, within the ligation mixture, after type II enzyme and DNA ligase treatment, target fragments will line up according to sticky end matching, forming linear concatemers, with a much higher abundance compared to the circularized form. The ability of *B. subtilis* to uptake these linear concatemers and circularize them inside cells significantly increases assembly capacity. Using OGAB, it was shown that 98 out of 100 examined colonies were correct when five gene cassettes were ligated with a linearized backbone.

2.2.6. SIRA

SIRA requires the recruitment of a serine integrase and a recombination directionality factor (RDF) to accomplish site-specific recombination. The serine integrase alone is able to cut DNA fragments at *attB* and *attP* sites in synchrony, and re-ligate them to form two new sites, *attL* and *attR*. However, the addition of RDF in assembly mixtures can completely reverse the reaction. To maintain efficient recombination, exact matching of 2-base pair overlaps of

adjacent fragments is required. Orthogonal recombination sites are necessary for ordered recombination of multiple fragments.

High flexibility is the main advantage of this method. DNA fragments can be deleted from the recombined vectors or substituted by one or more fragments. A two-step replacement strategy was developed to improve the efficiency of fragment substitution and addition. In the first step, a target gene was replaced by an antibiotic resistance cassette through serine integrase- and RDF-dependent *attL/attR* recombination. Then, with serine integrase alone, three DNA fragments with the appropriate flanking sites replaced the antibiotic resistance marker. SIRA was also utilized to construct metabolic pathways, alter gene order to adjust to balance expression levels, and optimize synthetic genetic pathways (Colloms et al., 2014).

3. Tools for Genetic Block Visualization

The development of computational biology and bioinformatic tools has immensely aided the modeling and prediction of a myriad of intricate biological mechanisms. From computational genomics (Koonin, 2001) to evolutionary biology (Torday, 2012), cutting-edge computational tools allow scientists an *in silico* first approach to the design of wet-lab experiments, lowering the possibility of introducing the rather inevitable human errors. This is precisely what has been achieved with the visualization tools for pathway assembly. Nowadays, one can virtually manipulate DNA parts as building blocks with a friendly user interface program, allowing the elaboration of complex gene circuits and metabolic pathways. This also opens the door for the accurate design of intricate combinatorial pathway constructs, permitting to precisely keep track of restriction sites, homology regions, as well as to seamlessly design primers. **Table 2** contains the current most popular computational tools for design and visualization of DNA

circuits. Here we will pay special attention to j5 since it is a free, open-source web software that anyone can access from a computer with an Internet connection.

Table 2: Computational tools for visualization and design of DNA constructs.

Software	Source	Reference
Vector NTI	Life Technology	(Lu and Moriyama, 2004)
j5	http://j5.jbei.org	(Hillson et al., 2012)
GeneDesigner	https://www.dna20.com/resources/genedesigner	(Villalobos et al., 2006)
GeneDesign	https://genedesign.org	(Richardson et al., 2006)
Gene Composer	http://www.genecomposer.net/	(Lorimer et al., 2009)

j5 DNA assembly software was first introduced in 2012 by Hillson *et al.* with the goal of automating the design of scar-less multipart constructs (Hillson et al., 2012). This software was designed with the capability of adopting the requirements of common assembly techniques like isothermal Gibson (Gibson et al., 2008) and Golden Gate (Engler et al., 2008). A great feature enabled in this software is the cost-analysis algorithm that compares the cost of synthesizing a DNA fragment to laborious PCR assembly of the same fragment. This software also contains algorithms to create precise oligonucleotides to provide homology regions for the case of SLIC (Li and Elledge, 2007), Gibson (Gibson et al., 2009), and CPEC (Quan and Tian, 2009) (**Table 1**). One of the criteria for this design relies on avoiding the creation of fragments that share high similarity in order to prevent unspecific assemblies. Moreover, for Golden Gate assembly, an algorithm built in the software allows the selection of appropriate overhang combinations to circumvent potential self-cohesive or off-target annealing events.

j5 and other software like GeneDesig (Richardson et al., 2006) and Vector NTI (Lu and Moriyama, 2004) do not focus on the biological activity of the DNA fragments to be assembled. As a visualization tool, these software solely aid in the assembly design of the

genetic constructs. Nevertheless, tools like the RBS calculator, are great complementary tools since their main purpose is to generate DNA variants of regulatory fragments with tailored biological functionalities. The junction of software like these, together with robotics and microfluidics, open the door for the implementation of combinatorial approaches for assembly of pathways to exploit biological diversity. The concept of combinatorial design is introduced in the following section.

4. Combinatorial Assembly

Pathway assembly represents an important step in the engineering of microbes towards the production of value-added compounds. However, after introducing the heterologous genes into the host, an ample room for optimization remains available. Combinatorial pathway assembly is a concept that has been recently coined as the product of the desire to exploit variability in the components of a pathway. For example, this variability, commonly introduced by some types of mutagenesis, can be targeted to create libraries of promoters with different strengths. As a result, a more efficient pathway may be generated due to an optimal balance in gene expression.

As it will be depicted, all of the combinatorial assembly techniques make use of the assembly methods mentioned in the previous section. It is in this case that the rapidness of assembly plays a very important role. Only the most efficient techniques, generally the ones that allow assembly with only a few steps, will permit the wide application of combinatorial design. Techniques limited by restriction digestion and multiple cloning steps may not be considered as an option when a combinatorial approach will be put in play. **Table 3** displays the predominant combinatorial techniques described in this section.

This section will describe several approaches to assemble pathways for the production of diverse metabolites following combinatorial techniques. Also, an extension to the combinatorial approach towards production of natural products will be included. The intrinsic modularity of the pathways for natural product production allows combinatorial design to be exploited at its fullest.

4.1. *In vivo* combinatorial assembly

4.1.1. Cocktail assembly and chromosomal integration of metabolic pathways

An interesting feature in the genomic DNA of *S. cerevisiae* is the presence of retrotransposon sequence denominated δ-site. This sequence has more than 100 copies spread out on the genome and has been used as recombination sites for the integration of heterologous genes in the yeast since 1990 (Sakai et al., 1990), and continues to be exploited for metabolic engineering studies (Lee and Da Silva, 1997; Shao et al., 2009; Yamada et al., 2010). In 2010, a technique denominated cocktail δ-integration was developed to introduce cellulose-degrading genes from the fungus *Trichoderma reesei* into the δ sites of *S. cerevisiae*. *T. reesei* is a fungus known for possessing various enzymes capable of degrading cellulose (Yamada et al., 2010). Depending on the environment, this fungus can balance the expression of such genes to efficiently degrade cellulose (Stricker et al., 2008). The dynamic nature of this process has not permitted to discern the appropriate expression levels of the genes required in this process.

Table 3: Techniques for combinatorial pathway design.

Combinatorial technique	Assembly foundation	Control level	Applications
δ-Cocktail Integration	Integration of genes into δ-sequence in yeast genome	- Transcriptional *(variability in integration events)*	- Integration of cellulolytic enzymes in *S. cerevisiae* (Yamada et al., 2010). - Introduction of five-gene isobutanol pathway and eight-gene amorpha-4,11-diene pathway in *S. cerevisiae* .
COMPACTER	DNA Assembler	- Transcriptional *(promoter variants)*	- Construction of an enhanced 3-gene xylose and a 2-gene cellobiose utilization pathway in *S. cerevisiae* controlled by a promoter library (Du et al., 2012).
Single-Step Linker-Based Assembly	Gibson Assembly	- Transcriptional *(promoter variants)* - Enzyme variability	- Restoring the acetate utilization pathway in *E. coli* (Ramon and Smith, 2011).
Single Strand Assembly		- Transcriptional *(promoter library)* - Translational *(RBS library)* - Enzyme variability	- Combinatorial assembly of variant promoters, RBS sequences, and reporter genes as proof of concept (Coussement et al., 2014).
MoClo	Golden Gate	- Transcriptional	- Creation of an 11-transcription unit construct conformed by 44 individual basic units (Weber et al., 2011a).
GB2.0		- Transcriptional - Translational	- Standardized modular assembly method for gene expression in plants (Sarrion-Perdigones et al., 2013).
ePathBrick	BioBrick	- Transcriptional - Translational	- Flavonoid production pathway in *E. coli* (Xu et al., 2012). - Engineering of the fatty acid central metabolism in *E. coli* leading to the highest reported productivity (Peng et al., 2013).

The goal of cocktail δ -integration was to develop a *S. cerevisiae* strain capable of degrading cellulose more efficiently without knowledge *a priori* of the required balanced expression levels of the genes. They incorporated three genes, namely, endoglucanase (EG), cellobiohydrolase (CBH), and β-glucosidase (BGL) with different auxotrophic markers into multiple δ sites of the yeast. After only two rounds of integration they obtained a strain with a higher activity of degrading phosphoric acid swollen cellulose. This increase was merely due to a balanced expression of the three genes, as opposed to a strain that simply contained the three genes overexpressed which had 13% less activity than their resulting variant with a balanced expression.

Recently, Yuan and Ching (Yuan and Ching, 2015) took the foundation of cocktail δ-integration to engineer *S. cerevisiae* for the production of two metabolites of interest in the biorenewables field, namely, isobutanol (a potential biofuel) and amorpha-4,11-diene (a precursor for artemisinin, the potent antimalarial medicine). They modified the δ-integration technique by introducing two coding sequences (CDS) per integration cassette driven by a bidirectional promoter. Also, each cassette contained a gene for phleomycin resistance, allowing them to control the integration frequency by varying the antibiotic concentration.

Proceeding with this idea they assembled 5 genes involved in isobutanol production into the yeast genomic DNA. After a single round of integration, they obtained a strain capable of producing 600 mg/L of isobutanol, an impressive increase, since the reference strain, containing the pathway in a multiple copy plasmid, was only able to produce around 6 mg/L. Following the same rationale, the mevalonate pathway consisting of 8 genes was integrated into the δ-sites with 100% efficiency (all randomly selected mutants contained the full

pathway). Although their final strain did not reach to the maximal level as reported in literature with a different yeast strain as the host (Newman et al., 2006; Tsuruta et al., 2009), they did consistently observe that the strain with higher amorpha-4,11-diene accumulation gave a low transcription abundance of the gene susceptible to cause feedback inhibition specific for this pathway. This allowed concluding that, indeed, their technique can go beyond what was originally shown in the DNA Assembler method for chromosome-based pathway assembly (Shao et al., 2009), and be utilized as a combinatorial way to fine tune the expression levels of individual genes to obtain more efficient phenotypes.

The combinatorial rationale in the aforementioned technique pertains to balancing the expression of the genes constituting the pathway. Work like this can help decipher the ways in which the hosts are naturally able to produce the metabolite of interest balance the expression of their genes. Nevertheless, it is important to keep in mind that, in most cases, metabolic pathways have to be tailored to a specific host. That is, it is common that a pathway resulting from a combinatorial approach does not yield the same results when transferred between strains even if they are from the same species. Du and Yuan *et al*. (Du et al., 2012) demonstrated this when they engineered an efficient xylose- and cellobiose-utilization pathways in strains of *S. cerevisiae* with different genetic backgrounds. Similar to the cocktail integration technique, this group targeted combinatorial expression levels in their constructed pathways. However, unlike having multiple integrations into the yeast chromosomes, they created libraries of mutant promoters with confirmed varying strengths obtained by nucleotide analog mutagenesis (Ueda et al., 1995).

The name of this combinatorial approach is COMPACTER, which stands for ‘customized optimization of metabolic pathways by combinatorial transcriptional engineering’ and it is

based on the DNA Assembler (Shao et al., 2009) technique described in the previous section. To analyze the colony variants, they developed a rigorous screening method that allowed them to select strains with higher pathway activity based on colony sizes. Subsequential refinement of the selected mutants helped rule out the possibility of background mutations and host adaptation to xylose or cellobiose. For each strain studied, they were able to obtain a mutant pathway that enabled optimal utilization of either xylose or cellobiose as sole carbon source.

Translational optimization in combinatorial pathway design adds a second level of control to enhance heterologous production of biofuels and chemicals. This approach includes, but is not limited to, modulating the ribosome binding site sequences (RBS) (Levin-Karp et al., 2013; Salis, 2011), combining enzyme homologues (Kim et al., 2013), or implementing libraries of mutant enzymes. The following example demonstrates how an efficient xylose utilization pathway was constructed by exploring thousands of combinations of enzyme homologues with various features. This approach, again, took advantage of the DNA Assembler (Shao et al., 2009) method discussed previously.

There are multiple microorganisms that are capable of utilizing xylose as a carbon source. This invaluable feature is enabled by the presence of three key enzymes, XR, XDH, and xylulose kinase (XK). Although this set of enzymes share a certain level of homology among xylose-utilizing microbes, they all display differences in cofactor selectivity. These differences were explored by Kim *et al.*(Kim et al., 2013) through assembling around 8,000 xylose-utilization pathway variants as a result of the combination of multiple enzyme homologues of the three enzymes originated from *Scheffersomyces stipitis*, a yeast well-known for its exceptional assimilation of xylose as a carbon source (Jeffries, 2006; Jeffries et al., 2007; Jeffries and Jin, 2004). Their pathway library consisted of 20 XR, 22 XDH, and 19 XK

homologues from various hosts, yielding a total of 8360 possible combinations. Also they tested the performance of the pathways in different *S. cerevisiae* strains, such as laboratory (InvSc1) and industrial (ATCC 4124) strains. Coupled with a screening method based on colony size, xylose consumption, and ethanol production, they were able to obtain combinations that allowed efficient xylose utilization compared to other highly efficient xylose-utilizing strains. It was observed that activities among the enzymes involved were, XKS>XDH>XR, represented by specific activities, which at least set a starting point for further understanding the optimal balance in expression level for this 3-gene pathway. They also proved that the performance of the pathway was strain dependent as well as carbon source dependent. Once again, it was proven that a combinatorial approach allows a fast way to exploit random combinations of pathway elements that yield better performances.

4.2. *In vitro* combinatorial assembly

The following examples show combinatorial pathway design based on *in vitro* assembly techniques. So far, we have shown how two levels of pathway assembly variability can be independently targeted by combinatorial techniques. Could it be possible to merge both into one to create an even larger pool of pathway variants?

In 2010 Ramon and Smith (Ramon and Smith, 2011) developed a technique founded upon isothermal assembly (Gibson) (Gibson et al., 2009), which targeted the combinatorial assembly of promoters and gene cassettes. The fragments involved in the construct were flanked with 40-bp unique linker sequences to drive the assembly in an orderly manner. This opened the combination at two control levels: one referring to the promoter strength and the other to the variable features displayed by homologue genes. The combinatorial assembly was done *in vitro*

and then transferred to the host organism, unlike the previous examples, in which the library of pathways was directly assembled and screened in the same organism. As an example to prove the functionality of their method, they restored the acetate assimilation pathway, consisting of the genes acetate kinase (*ackA*) and phosphotransacetylase (*pta*), into a mutant *E. coli* strain incapable of growing on acetate as sole carbon source. In total, they selected three different promoters and four homologues for each enzyme to yield a total of 144 combinations. They were able to select positive transformants from the plates containing acetate as sole carbon source. However, the distribution analysis of their mutants showed misrepresentation of some combinations that were later attributed to uneven concentrations of individual cassettes before the assembly. Predominance of some combinations could also be originated from better functioning pathways; showing again that combinatorial assembly can function as a tool to efficiently find promoters and gene homologues that display better performance in a multigenic pathway.

Recently, Coussement *et al*. (Coussement et al., 2014) showed that targeting all three levels of pathway control in *E. coli* was feasible using fluorescent reporter gene as a proof of concept. The library of assembled modules contained variability in promoters, RBS sequences, and CDS. To enable this combinatorial approach, they relied on Gibson Assembly (Gibson et al., 2009) of oligonucleotides encoding different regions of the construct. They named this approach as the single-stranded assembly technique (SSA). Since they targeted variability in all three elements of their target, they devised three ways in which their oligonucleotides could be assembled, and tested for the one yielding the highest efficiency. In their first case, a single-very long oligonucleotide contained all three sections namely, promoter, RBS, and CDS. Their second scheme partitioned the sequence in three smaller oligos, while their last scheme

partitioned the sequence into two oligos of similar sizes. Their assembly analysis showed that the third scheme had the highest efficiency followed by the second scheme with almost 100% efficiency, while the first scheme had only around 25% efficiency.

When implementing the third scheme into the combinatorial assembly with varying promoter and RBS, they were able to observe that the levels of fluorescence resulting from these combinations exceeded the values obtained when just the promoter or the RBS were individually subjected to variability. Finally, when incorporating two reporter genes in control of a promoter library, they observed good orthogonality in the expression of the genes. They were able to conclude that the method allows to independently and in parallel tune the transcript levels of the genes. However, even though their best scheme showed excellent assembly efficiencies, longer oligonucleotides may be required to assemble longer pathways, thus limiting the potential of this technique to small pathways, or to multiple steps of assembly, a common drawback of *in vitro* assembly techniques.

Further, *in vitro* combinatorial techniques have originated from well-established assembly techniques. For example, the Golden Gate Assembly technique has allowed the origination of two combinatorial assembly techniques, namely MoClo (Weber et al., 2011a) and GoldenBraid (Sarrion-Perdigones et al., 2011). The main aim of these techniques is to standardize the way combinatorial pathways can be assembled. In MoClo and GoldenBraid, they subdivided all the genetic elements that will confer tiers or levels on their DNA composite, starting from level zero, which consists of individual genetic elements that can include, but are not limited to RBS, CDS, terminators, and signal peptides. In a series of one-pot assembly reactions, they combine the level-zero parts to form a level-one composite. Successively, the level-one constructs can be combined to form level-two assemblies. This is enabled by flanking each module with

fusion sites and type II restriction sites (*Bsa*I and *Bpi*I); destination vectors allowed for different selection methods of positive transformants. In 2013, an update to GoldenBraid was published and was denominated GoldenBraid 2.0 (GB2.0) (Sarrion-Perdigones et al., 2013). Similar to the two previous examples, the inventors of GB2.0 made an analogy of their technique to the foundation English grammar. Their level-zero parts function just like morphemes and grammemes as units in linguistics that allow the construction words to complete phrases. Mainly, GB 2.0 was developed as a standardized procedure to assemble pathways for plant metabolic engineering with combinatorial applicability. As an updated version of GoldenBraid, GB2.0 incorporates several new tools that simplify the overall assembly process. First, they introduced a *BtgZ*I site into the entry vector. The ability of this enzyme to cut 10 nucleotides from its recognition site enables the capability of a "dual-release" option for each GBpart, speeding up the construction of multigenic assemblies. Also they have arranged a collection of standard parts compatible with their technique available as an open source, as well as web applications that allow a friendly interface for *in silico* pathway design. The authors envisioned their technique as a revolutionary method that will allow fine automation of the assembly process. They claimed that the use of their technique will potentialize the advancement in, not only plant synthetic biology, but also for microorganisms, since their method facilitates exchangeability of genetic parts.

Natural products represent the perfect example for implementing a combinatorial approach to enable exploration of the vast pool of metabolite variants. Although the discovery of natural products is a topic of great importance, the intricate regulations embedded in gene clusters involved in the production of such secondary metabolites hamper quick advancement in this area. New DNA technologies have to be implemented to ease drug discovery and one of them

is fast combinatorial ways to assemble gene clusters or independent enzymes involved in secondary metabolism.

A classic example of combinatorial assembly for natural products is the scrambling of enzymes belonging to the family of type I polyketide synthase (PKS). The enzymes involved in this family resemble an industrial assembly line in which each module carries out a specific reaction. This feature allows to easily apply the concept of combinatorial design, in which each module can be removed, mutated, or substituted by putative enzymes to finally yield the desired product or even create unnatural ones. Menzella *et al.* (Menzella et al., 2005) developed a combinatorial approach to assemble codon-optimized PKS modules to produce unnatural triketide lactones in *E. coli.* They designed 11 donor and 14 acceptor modules from which 72 out of the 154 possible combinations were able to produce a potential unnatural triketide lactone. In addition, in 2007 Katsuyama *et al*. (Katsuyama et al., 2007) targeted type III PKS and produced a series of metabolites from the stilbene and flavonoid family by combining enzymes from different non-microbial sources. By mixing enzymes involved in precursor formation and post-PKS tailoring, the authors enabled the diversification of the originally targeted products and discivered 36 new secondary metabolites.

Most of the research in this field has been done with the aid of classic DNA cloning methodologies, in which time-consuming steps limited the research to case-by-case scenarios. The development of new assembly techniques and synthetic biology tools, hence, represents a great potential for new natural product discovery. Even though there are only a few examples of applying such techniques in this area, it is not hard to foresee the great potential that they possess. For example, a recently developed synthetic biology strategy based on a plug-and-play scaffold allowed for the heterologous production of spectinabilin in *Streptomyces lividans*

(Shao et al., 2013). This technique allowed the bypass of endogenous regulation by setting the expression of the genes in the cluster to constitutive promoters. Founded upon the DNA Assembler (Shao et al., 2009), this technique has also allowed the production of polycyclic tetramate macrolactams and the discovery of new derivatives from the same family (Luo et al., 2013).

The methods exemplified in this text have demonstrated vast improvements in terms of standardization, efficacy, and versatility. Nowadays the limitations do not solely rely on the assembly method of choice but rather on developing fast screening techniques that permit to easily explore the multiple phenotypes generated along the assembly process. Hence, to leverage rapid strain engineering enabled by combinatorial pathway assembly, high-throughput screening techniques must be readily available.

5. Bibliography

Anderson, J. C., Dueber, J., Leguia, M., Wu, G., Goler, J., Arkin, A., Keasling, J., 2010. BglBricks: A flexible standard for biological part assembly. Journal of Biological Engineering. 4**,** 1.

Annaluru, N., Muller, H. l. s., Mitchell, L. A., 2014. Total synthesis of a functional designer eukaryotic chromosome. Science. 344**,** 55.

Bitinaite, J., Rubino, M., Varma, K. H., Schildkraut, I., Vaisvila, R., Vaiskunaite, R., 2007. USER™ friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Research. 35**,** 1992-2002.

Blake, W. J., Chapman, B. A., Zindal, A., Lee, M. E., Lippow, S. M., Baynes, B. M., 2010. Pairwise selection assembly for sequence-independent construction of long-length DNA. Nucleic Acids Research. 38**,** 2594-602.

Bloch, S. E., Schmidt-Dannert, C., 2014. Construction of a Chimeric Biosynthetic Pathway for the De Novo Biosynthesis of Rosmarinic Acid in *Escherichia coli*. ChemBioChem.

Boratyn Grzegorz, M., Schäffer Alejandro, A., Agarwala, R., Altschul Stephen, F., Lipman David, J., Madden Thomas, L., 2012. Domain enhanced lookup time accelerated BLAST. Biology Direct. 7**,** 12.

Cao, P., Wang, L., Zhou, G., Wang, Y., Chen, Y., 2014. Rapid assembly of multiple DNA fragments through direct transformation of PCR products into *E. coli* and *Lactobacillus*. Plasmid. 76**,** 40-46.

Chen, W.-H., Qin, Z.-J., Wang, J., Zhao, G.-P., 2013. The MASTER (methylation- assisted tailorable ends rational) ligation method for seamless DNA assembly. Nucleic Acids Research. 41**,** e93.

Clementina, D., James, M. C., Elliot, N. M., Ramon, G., 2011. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. Nature. 476**,** 355.

Cobb, R. E., Ning, J. C., Zhao, H., 2014. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. Journal of Industrial Microbiology and Biotechnology. 41**,** 469-77.

Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A., Keasling, J. D., Rosser, S. J., 2014. Rapid metabolic pathway assembly and

modification using serine integrase site-specific recombination. Nucleic Acids Research. 42**,** e23.

Coussement, P., Maertens, J., Beauprez, J., Van Bellegem, W., De Mey, M., 2014. One step DNA assembly for combinatorial metabolic engineering. Metabolic Engineering. 23**,** 70-77.

de Kok, S., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., Platt, D., Shapland, E. B., Serber, Z., Dean, J., Newman, J. D., Chandran, S. S., 2014. Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synthetic Biology. 3**,** 97.

Du, J., Yuan, Y., Si, T., Lian, J., Zhao, H., 2012. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. Nucleic Acids Research. 40**,** e142.

Engler, C., Gruetzner, R., Kandzia, R., Marillonnet, S., 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes (One- Pot DNA Shuffling Method). PLoS ONE. 4**,** e5553.

Engler, C., Kandzia, R., Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability (High Throughput Cloning Method). PLoS ONE. 3**,** e3647.

Gao, S., Han, L., •, L. Z., •, M. G., •, S. Y., •, Y. J., Chen, D., One-step integration of multiple genes into the oleaginous yeast *Yarrowia lipolytica*. Biotechnology Letters, 2014.

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., Smith, H. O., 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science. 319**,** 1215.

Gibson, D. G., Smith, H. O., Hutchison, C. A., III, Venter, J. C., Merryman, C., 2010. Chemical synthesis of the mouse mitochondrial genome. Nature Methods. 7**,** 901.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., III, Smith, H. O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods. 6**,** 343.

Guye, P., Li, Y., Wroblewska, L., Duportet, X., Weiss, R., 2013. Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic Acids Research. 41**,** e156.

Hillson, N. J., Rosengarten, R. D., Keasling, J. D., 2012. j5 DNA assembly design automation software. ACS Synthetic Biology. 1**,** 14.

Jeffries, T. W., 2006. Engineering yeasts for xylose metabolism. Curr Opin Biotechnol. 17**,** 320-6.

Jeffries, T. W., Grigoriev, I. V., Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y. S., Passoth, V., Richardson, P. M., 2007. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. Nat Biotechnol. 25**,** 319-26.

Jeffries, T. W., Jin, Y. S., 2004. Metabolic engineering for improved fermentation of pentoses by yeasts. Appl Microbiol Biotechnol. 63**,** 495-509.

Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandeau-Nelson, M. D., Reilly, P. J., 2011. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochemistry. 12**,** 44.

Kahl, L. J., Endy, D., 2013. A survey of enabling technologies in synthetic biology. Journal of biological engineering. 7**,** 13.

Katsuyama, Y., Funa, N., Miyahisa, I., Horinouchi, S., 2007. Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. Chem Biol. 14**,** 613-21.

Kim, B., Du, J., Eriksen, D. T., Zhao, H., 2013. Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. Appl Environ Microbiol. 79**,** 931-41.

Koonin, E. V., 2001. Computational genomics. Current Biology. 11**,** R155-R158.

Lane, M. D., Seelig, B., 2014. Advances in the directed evolution of proteins. Current Opinion in Chemical Biology. 22**,** 129-136.

Lee, F. W., Da Silva, N. A., 1997. Sequential delta-integration for the regulated insertion of cloned genes in *Saccharomyces cerevisiae*. Biotechnol Prog. 13**,** 368-73.

Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., Keasling, J. D., 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Current Opinion in Biotechnology. 19**,** 556-563.

Levin-Karp, A., Barenholz, U., Bareia, T., Dayagi, M., Zelcbuch, L., Antonovsky, N., Noor, E., Milo, R., 2013. Quantifying translational coupling in *E. coli* synthetic operons using RBS modulation and fluorescent reporters. ACS synthetic biology. 2**,** 327.

Li, M. Z., Elledge, S. J., 2005. MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. Nature Genetics. 37**,** 311.

Li, M. Z., Elledge, S. J., 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nature Methods. 4**,** 251.

Li, Y., Cirino, P. C., Recent advances in engineering proteins for biocatalysis. Vol. 111, 2014, pp. 1273-1287.

Lorimer, D., Raymond, A., Walchli, J., Mixon, M., Barrow, A., Wallace, E., Grice, R., Burgin, A., Stewart, L., Gene Composer: database software for protein construct design, codon engineering, and gene synthesis. Vol. 9, 2009, pp. 36-36.

Lu, G., Moriyama, E. N., 2004. Vector NTI, a balanced all-in-one sequence analysis suite. Briefings in bioinformatics. 5**,** 378.

Luo, Y., Huang, H., Liang, J., Wang, M., Lu, L., Shao, Z., Cobb, R. E., Zhao, H., 2013. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. Nature communications. 4**,** 2894.

Ma, H., Kunes, S., Schatz, P. J., Botstein, D., 1987. Plasmid construction by homologous recombination in yeast. Gene. 58**,** 201-216.

Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., Santi, D. V., 2005. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat Biotechnol. 23**,** 1171-6.

Natalay, K., Vladimir, L., 2006. TAR cloning: insights into gene function, long- range haplotypes and genome structure and evolution. Nature Reviews Genetics. 7**,** 805.

Newman, J. D., Marshall, J., Chang, M., Nowroozi, F., Paradise, E., Pitera, D., Newman, K. L., Keasling, J. D., 2006. High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. Biotechnol Bioeng. 95**,** 684-91.

Pandya, C., Farelli, J. D., Dunaway-Mariano, D., Allen, K. N., 2014. Enzyme Promiscuity: Engine of Evolutionary Innovation. Journal of Biological Chemistry. R114.572990.

Peng, X., Qin, G., Wenya, W., Lynn, W., Adam, G. W. B., Cynthia, H. C., Mattheos, A. G. K., 2013. Modular optimization of multi- gene pathways for fatty acids production in *E. coli*. Nature Communications. 4**,** 1409.

Priscilla, E. M. P., Ron, W., 2009. The second wave of synthetic biology: from modules to systems. Nature Reviews Molecular Cell Biology. 10**,** 410.

Quan, J., Tian, J., 2009. Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways (CPEC Cloning). PLoS ONE. 4**,** e6441.

Rabinovitch-Deere, C. A., Rodriguez, G. M., Oliver, J. W. K., Atsumi, S., 2013. Synthetic biology and metabolic engineering approaches to produce biofuels. Chemical Reviews. 113**,** 4611-4632.

Ramon, A., Smith, H., 2011. Single- step linker- based combinatorial assembly of promoter and gene cassettes for pathway engineering. Biotechnol Lett. 33**,** 549-555.

Redden, H., Morse, N., Alper, H. S., 2014. The synthetic biology toolbox for tuning gene expression in yeast. FEMS Yeast Res.

Ren, Y., Perepelov, A. V., Wang, H., Zhang, H., Knirel, Y. A., Wang, L., Chen, W., 2010. Biochemical characterization of GDP- l-fucose de novo synthesis pathway in fungus *Mortierella alpina*. Biochemical and Biophysical Research Communications. 391**,** 1663-1669.

Richardson, S. M., Wheelan, S. J., Yarrington, R. M., Boeke, J. D., 2006. GeneDesign: rapid, automated design of multikilobase synthetic genes. Genome research. 16**,** 550.

Robert, C., 2009. The changing economics of DNA synthesis. Nature Biotechnology. 27**,** 1091.

Royce, L. A., Liu, P., Stebbins, M. J., Hanson, B. C., Jarboe, L. R., 2013. The damaging effects of short chain fatty acids on *Escherichia coli* membranes. Applied microbiology and biotechnology. 97**,** 8317.

Sakai, A., Shimizu, Y., Hishinuma, F., 1990. Integration of heterologous genes into the chromosome of *Saccharomyces cerevisiae* using a delta sequence of yeast retrotransposon Ty. Appl Microbiol Biotechnol. 33**,** 302-306.

Salis, H. M., 2011. The ribosome binding site calculator. Methods Enzymol. 498**,** 19-42.

Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-del-Carmen, A., Granell, A., Orzaez, D., Peccoud, J., 2011. GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. PLoS ONE. 6.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., Orzaez, D., 2013. GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. Plant Physiology. 162**,** 1618.

Shao, Z., Luo, Y., Zhao, H., 2011. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. Mol. BioSyst. 7**,** 1056-1059.

Shao, Z., Luo, Y., Zhao, H., DNA Assembler Method for Construction of Zeaxanthin-Producing Strains of *Saccharomyces cerevisiae*. Microbial Carotenoids From Fungi, Chapter 17, 2012, pp. 251-262.

Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., Zhao, H., 2013. Refactoring the silent spectinabilin gene cluster using a plug-and- play scaffold. ACS synthetic biology. 2**,** 662.

Shao, Z., Zhao, H., Zhao, H., 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic acids research. 37**,** e16.

Stevenson, J., Krycer, J. R., Phan, L., Brown, A. J., 2013. A practical comparison of ligation-independent cloning techniques. PloS one. 8**,** e83888.

Stricker, A. R., Mach, R. L., de Graaff, L. H., 2008. Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). Appl Microbiol Biotechnol. 78**,** 211-20.

Torday, J. S., 2012. Evolutionary Biology. John Wiley & Sons, Inc. US.

Torella, J. P., Lienert, F., Boehm, C. R., Chen, J.-H., Way, J. C., Silver, P. A., 2014. Unique nucleotide sequence- guided assembly of repetitive DNA parts for synthetic biology applications. Nature Protocols. 9**,** 2075.

Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., French, C. E., 2014. PaperClip: rapid multi-part DNA assembly from existing libraries. Nucleic Acids Res. 42**,** e154.

Tsuge, K., Matsui, K., Itaya, M., 2003. One step assembly of multiple DNA fragments with a designed order and orientation in *Bacillus subtilis* plasmid. Nucleic acids research. 31**,** e133.

Tsuruta, H., Paddon, C. J., Eng, D., Lenihan, J. R., Horning, T., Anthony, L. C., Regentin, R., Keasling, J. D., Renninger, N. S., Newman, J. D., 2009. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. PLoS One. 4**,** e4489.

Ueda, H., Nakanishi, M., Suzuki, E., Nagamune, T., 1995. Enhancement of mutation frequency with nucleotide triphosphate analogs in PCR random mutagenesis. Journal of Fermentation and Bioengineering. 79**,** 303-305.

Villalobos, A., Ness, J., Gustafsson, C., Minshull, J., Govindarajan, S., 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics. 7**,** 285.

Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S., Peccoud, J., 2011a. A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLoS ONE. 6.

Weber, E., Gruetzner, R., Werner, S., Engler, C., Marillonnet, S., 2011b. Assembly of designer TAL effectors by Golden Gate cloning. PloS one. 6**,** e19722.

Xiao, H., Shao, Z. Y., Jiang, Y., Dole, S., Zhao, H. M., 2014. Exploiting Issatchenkia orientalis SD108 for succinic acid production. Microbial Cell Factories. 13.

Xu, P., Vansiri, A., Bhan, N., Koffas, M. A. G., 2012. ePathBrick: a synthetic biology platform for engineering metabolic pathways in E. coli. ACS synthetic biology. 1**,** 256.

Yamada, R., Taniguchi, N., Tanaka, T., Ogino, C., Fukuda, H., Kondo, A., 2010. Cocktail delta- integration: a novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. Microbial cell factories. 9**,** 32.

Yuan, J., Ching, C. B., 2015. Combinatorial assembly of large biochemical pathways into yeast chromosomes for improved production of value-added compounds. ACS Synth Biol. 4**,** 23-31.

Zhang, Y., Werling, U., Edelmann, W., 2012. SLiCE: a novel bacterial cell extract- based DNA cloning method. Nucleic Acids Research. 40**,** e55-e55.